



(22) Date de dépôt/Filing Date: 1997/03/19

(41) Mise à la disp. pub./Open to Public Insp.: 1997/09/19

(45) Date de délivrance/Issue Date: 2012/11/13

(62) Demande originale/Original Application: 2 458 872

(30) Priorités/Priorities: 1996/03/19 (JP63410/1996);
1997/03/05 (JP69163/1997)

(51) Cl.Int./Int.Cl. *C12N 15/18* (2006.01),
C07K 14/475 (2006.01), *C07K 14/485* (2006.01),
C07K 16/22 (2006.01), *C12N 15/12* (2006.01),
C07K 14/705 (2006.01)

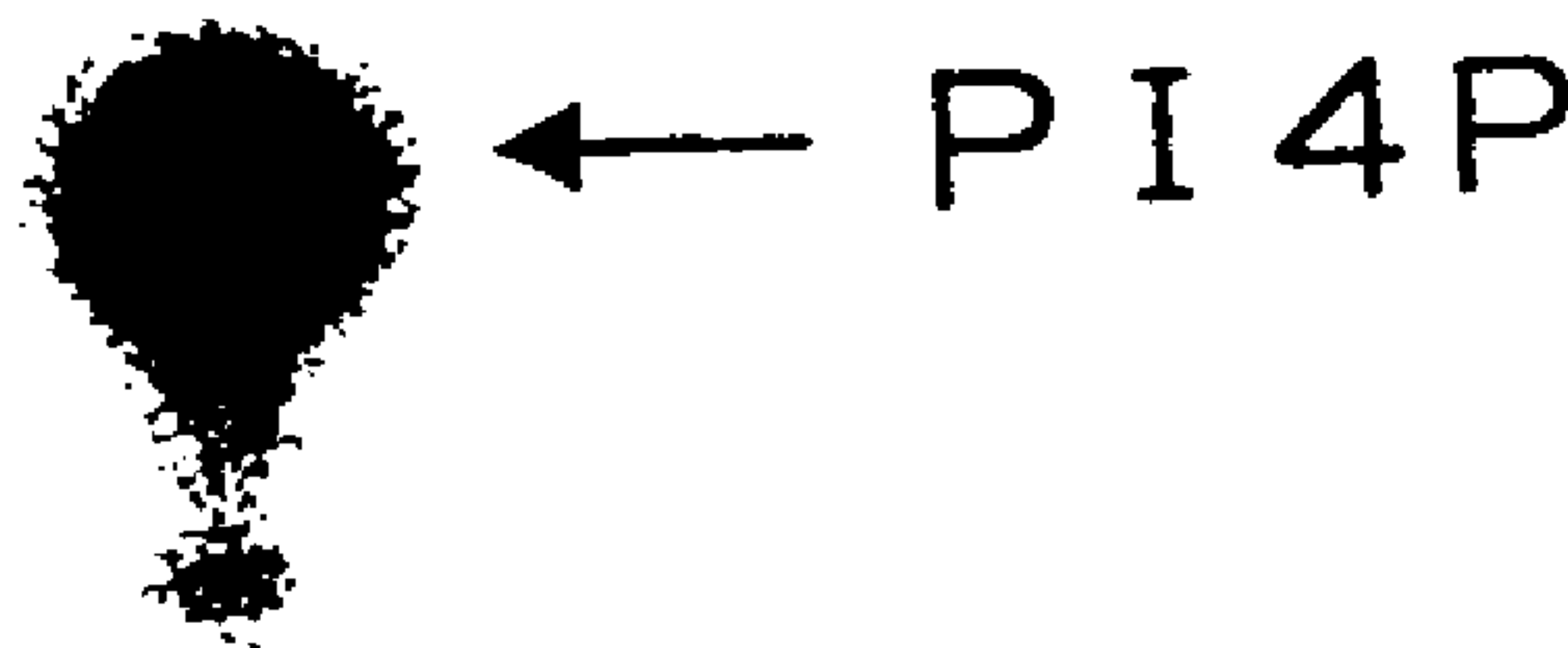
(72) Inventeur/Inventor:
HORIE, MASATO, JP

(73) Propriétaire/Owner:
OTSUKA PHARMACEUTICAL CO., LTD., JP

(74) Agent: MARKS & CLERK

(54) Titre : PROTEINES DE TYPE NEL DE TYPE 1 ET DE TYPE 2 ET POLYNUCLEOTIDES LES CODANT

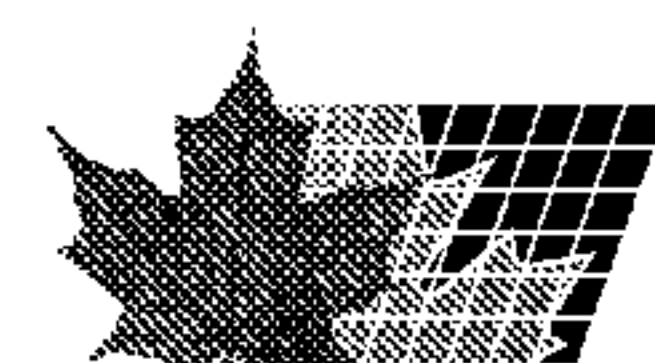
(54) Title: NEL-RELATED PROTEINS TYPE 1 AND 2, AND POLYNUCLEOTIDES ENCODING THEM



1 2

(57) Abrégé/Abstract:

Disclosed is a pharmaceutical composition for treating prostate cancer, the pharmaceutical composition comprising an antibody against human TMP-2 protein together with a pharmaceutically acceptable carrier. In one embodiment, the antibody is conjugated to a therapeutic moiety which can be a cytotoxic agent or a radioisotope.



ABSTRACT

Disclosed is a pharmaceutical composition for treating prostate cancer, the pharmaceutical composition comprising an antibody against human TMP-2 protein together with a pharmaceutically acceptable carrier. In one embodiment, the antibody is conjugated to a therapeutic moiety which can be a cytotoxic agent or a radioisotope.

DEMANDES OU BREVETS VOLUMINEUX

**LA PRÉSENTE PARTIE DE CETTE DEMANDE OU CE BREVETS
COMPREND PLUS D'UN TOME.**

CECI EST LE TOME __1__ DE __2__

NOTE: Pour les tomes additionels, veuillez contacter le Bureau Canadien des Brevets.

JUMBO APPLICATIONS / PATENTS

**THIS SECTION OF THE APPLICATION / PATENT CONTAINS MORE
THAN ONE VOLUME.**

THIS IS VOLUME __1__ OF __2__

NOTE: For additional volumes please contact the Canadian Patent Office.

-1-

**NEL-RELATED PROTEINS TYPE 1 AND 2,
AND POLYNUCLEOTIDES ENCODING THEM**

This is a divisional application of Canadian
5 Patent Application Serial No. 2,458,872 filed on March 23,
2004.

TECHNICAL FIELD

The present invention relates to a gene useful as
an indicator in the prophylaxis, diagnosis and treatment of
10 diseases in humans. More particularly, it relates to a
novel human gene analogous to rat, mouse, yeast, nematode
and known human genes, among others, and utilizable, after
cDNA analysis thereof, chromosome mapping of cDNA and
function analysis of cDNA, in gene diagnosis using said gene
15 and in developing a novel therapeutic method. It should be
understood that the expression "the invention" and the like
encompasses the subject matter of both the parent and the
divisional applications.

BACKGROUND ART

20 The genetic information of a living thing has been
accumulated as sequences (DNA) of four bases, namely A, C, G
and T, which exist in cell nuclei. Said genetic information
has been preserved for line preservation and ontogeny of
each individual living thing.

25 In the case of human being, the number of said bases
is said to be about 3 billion (3×10^9) and supposedly there
are 50 to 100 thousand genes therein. Such genetic information
serves to maintain biological phenomena in that regulatory
proteins, structural proteins and enzymes are produced via
30 such route that mRNA is transcribed from a gene (DNA) and then
translated into a protein. Abnormalities in said route from

-2-

gene to protein translation are considered to be causative of abnormalities of life supporting systems, for example in cell proliferation and differentiation, hence causative of various diseases.

5 As a result of gene analyses so far made, a number of genes which may be expected to serve as useful materials in drug development, have been found, for example genes for various receptors such as insulin receptor and LDL receptor, genes involved in cell
10 proliferation and differentiation and genes for metabolic enzymes such as proteases, ATPase and superoxide dismutases.

 However, analysis of human genes and studies of the functions of the genes analyzed and of the relations
15 between the genes analyzed and various diseases have been just begun and many points remain unknown. Further analysis of novel genes, analysis of the functions thereof, studies of the relations between the genes analyzed and diseases, and studies for applying the genes
20 analyzed to gene diagnosis or for medicinal purposes, for instance, are therefore desired in the relevant art.

 If such a novel human gene as mentioned above can be provided, it will be possible to analyze the level of expression thereof in each cell and the structure and
25 function thereof and, through expression product analysis

-3-

and other studies, it may become possible to reveal the pathogenesis of a disease associated therewith, for example a genopathy or cancer, or diagnose and treat said disease, for instance. It is an object of the present invention to provide such a novel human gene.

For attaining the above object, the present inventors made intensive investigations and obtained the findings mentioned below. Based thereon, the present invention has now been completed.

10

DISCLOSURE OF INVENTION

Thus, the present inventors synthesized cDNAs based on mRNAs extracted from various tissues, inclusive of human fetal brain, adult blood vessels and placenta, constructed libraries by inserting them into vectors, allowing colonies of Escherichia coli transformed with said libraries to form on agar medium, picked up colonies at random and transferred to 96-well micro plates and registered a large number of human gene-containing E. coli clones.

20

Each clone thus registered was cultivated on a small size, DNA was extracted and purified, the four base-specifically terminating extension reactions were carried out by the dideoxy chain terminator method using the cDNA extracted as a template, and the base sequence of the gene was determined over about 400 bases from the

25

5' terminus thereof using an automatic DNA sequencer. Based on the thus-obtained base sequence information, a novel family gene analogous to known genes of animal and plant species such as bacteria, yeasts, nematodes, mice and humans was searched for.

The method of the above-mentioned cDNA analysis is detailedly described in the literature by Fujiwara, one of the present inventors [Fujiwara, Tsutomu, Saibo Kogaku (Cell Engineering), 14, 645-654 (1995)].

Among this group, there are novel receptors, DNA binding domain-containing transcription regulating factors, signal transmission system factors, metabolic enzymes and so forth. Based on the homology of the novel gene of the present invention as obtained by gene analysis to the genes analogous thereto, the product of the gene, hence the function of the protein, can approximately be estimated by analogy. Furthermore, such functions as enzyme activity and binding ability can be investigated by inserting the candidate gene into an expression vector to give a recombinant.

The present invention provides a pharmaceutical composition for treating prostate cancer, the pharmaceutical composition comprising an antibody against human TMP-2 protein together with a pharmaceutically acceptable carrier. In one embodiment, the antibody is conjugated to a therapeutic moiety which can be a cytotoxic agent or a radioisotope.

According to the present invention, there are provided a novel human gene characterized by containing a nucleotide sequence coding for an amino acid sequence defined by SEQ ID NO:1, :4, :7, :10, :13, :16, :19, :22, :25, :28, :31, :34, :37 or 40, a human gene characterized

-5-

by containing the nucleotide sequence defined by SEQ ID
NO:2, :5, :8, :11, :14, :17, :20, :23, :26, :29, :32,
:35, :38 or :41, respectively coding for the amino acid
sequence mentioned above, and a novel human gene
5 characterized by the nucleotide sequence defined by SEQ
ID NO:3, :6, :9, :12, :15, :18, :21, :24, :27, :30, :33,
:36, :39 or :42.

The symbols used herein for indicating amino
acids, peptides, nucleotides, nucleotide sequences and so
10 on are those recommended by IUPAC and IUB or in "Guide-
line for drafting specifications etc. including
nucleotide sequences or amino acid sequences" (edited by
the Japanese Patent Office), or those in conventional use
in the relevant field of art.

15 As specific examples of such gene of the
present invention, there may be mentioned genes deducible
from the DNA sequences of the clones designated as "GEN-
501D08", "GEN-080G01", "GEN-025F07", "GEN-076C09", "GEN-
331G07", "GEN-163D09", "GEN-078D05TA13", "GEN-423A12",
20 "GEN-092E10", "GEN-428B12", "GEN-073E07", "GEN-093E05"
and "GEN-077A09" shown later herein in Examples 1 to 11.
The respective nucleotide sequences are as shown in the
sequence listing.

These clones have an open reading frame
25 comprising nucleotides (nucleic acid) respectively coding

-6-

for the amino acids shown in the sequence listing. Their molecular weights were calculated at the values shown later herein in the respective examples. Hereinafter, these human genes of the present invention are sometimes
5 referred to as the designation used in Examples 1 to 11.

In the following, the human gene of the present invention is described in further detail.

As mentioned above, each human gene of the present invention is analogous to rat, mouse, yeast,
10 nematode and known human genes, among others, and can be utilized in human gene analysis based on the information about the genes analogous thereto and in studying the function of the gene analyzed and the relation between the gene analyzed and a disease. It is possible to use
15 said gene in gene diagnosis of the disease associated therewith and in exploitation studies of said gene for medicinal purposes.

The gene of the present invention is represented in terms of a single-stranded DNA sequence,
20 as shown under SEQ ID NO:2. It is to be noted, however, that the present invention also includes a DNA sequence complementary to such a single-stranded DNA sequence and a component comprising both. The sequence of the gene of the present invention as shown under SEQ ID NO:3n - 1
25 (where n is an integer of 1 to 14) is merely an example

-7-

of the codon combination encoding the respective amino acid residues. The gene of the present invention is not limited thereto but can of course have a DNA sequence in which the codons are arbitrarily selected and combined
5 for the respective amino acid residues. The codon selection can be made in the conventional manner, for example taking into consideration the codon utilization frequencies in the host to be used [Nucl. Acids Res., 9, 43-74 (1981)].

10 The gene of the present invention further includes DNA sequences coding for functional equivalents derived from the amino acid sequence mentioned above by partial amino acid or amino acid sequence substitution, deletion or addition. These polypeptides may be produced
15 by spontaneous modification (mutation) or may be obtained by posttranslational modification or by modifying the natural gene (of the present invention) by a technique of genetic engineering, for example by site-specific mutagenesis [Methods in Enzymology, 154, p. 350, 367-382
20 (1987); ibid., 100, p. 468 (1983); Nucleic Acids Research, 12, p. 9441 (1984); Zoku Seikagaku Jikken Koza (Sequel to Experiments in Biochemistry) 1, "Idensi Kenkyu-ho (Methods in Gene Research) II", edited by the Japan Biochemical Society, p. 105 (1986)] or synthesizing
25 mutant DNAs by a chemical synthetic technique such as the

-8-

phosphotriester method or phosphoamidite method [J. Am. Chem. Soc., 89, p. 4801 (1967); ibid., 91, p. 3350 (1969); Science, 150, p. 178 (1968); Tetrahedron Lett., 22, p. 1859 (1981); ibid., 24, p. 245 (1983)], or by
5 utilizing the techniques mentioned above in combination.

The protein encoded by the gene of the present invention can be expressed readily and stably by utilizing said gene, for example inserting it into a vector for use with a microorganism and cultivating the
10 microorganism thus transformed.

The protein obtained by utilizing the gene of the present invention can be used in specific antibody production. In this case, the protein producible in large quantities by the genetic engineering technique
15 mentioned above can be used as the component to serve as an antigen. The antibody obtained may be polyclonal or monoclonal and can be advantageously used in the purification, assay, discrimination or identification of the corresponding protein.

20 The gene of the present invention can be readily produced based on the sequence information thereof disclosed herein by using general genetic engineering techniques [cf. e.g. Molecular Cloning, 2nd Ed., Cold Spring Harbor Laboratory Press (1989); Zoku
25 Seikagaku Jikken Koza, "Idenshi Kenkyu-ho I, II and III",

-9-

edited by the Japan Biochemical Society (1986)].

This can be achieved, for example, by selecting a desired clone from a human cDNA library (prepared in the conventional manner from appropriate cells of origin in which the gene is expressed) using a probe or antibody specific to the gene of the present invention [e.g. Proc. Natl. Acad. Sci. USA, 78, 6613 (1981); Science, 222, 778 (1983)].

The cells of origin to be used in the above method are, for example, cells or tissues in which the gene in question is expressed, or cultured cells derived therefrom. Separation of total RNA, separation and purification of mRNA, conversion to (synthesis of) cDNA, cloning thereof and so on can be carried out by conventional methods. cDNA libraries are also commercially available and such cDNA libraries, for example various cDNA libraries available from Clontech Lab. Inc. can also be used in the above method.

Screening of the gene of the present invention from these cDNA libraries can be carried out by the conventional method mentioned above. These screening methods include, for example, the method comprising selecting a cDNA clone by immunological screening using an antibody specific to the protein produced by the corresponding cDNA, the technique of plaque or colony

-10-

hybridization using probes selectively binding to the desired DNA sequence, or a combination of these. As regards the probe to be used here, a DNA sequence chemically synthesized based on the information about the DNA sequence of the present invention is generally used. It is of course possible to use the gene of the present invention or fragments thereof as the probe.

Furthermore, a sense primer and an antisense primer designed based on the information about the partial amino acid sequence of a natural extract isolated and purified from cells or a tissue can be used as probes for screening.

For obtaining the gene of the present invention, the technique of DNA/RNA amplification by the PCR method [Science, 230, 1350-1354 (1984)] can suitably be employed. Particularly when the full-length cDNA can hardly be obtained from the library, the RACE method (rapid amplification of cDNA ends; Jikken Igaku (Experimental Medicine), 12 (6), 35-38 (1994)], in particular the 5'RACE method [Frohman, M. A., et al., Proc. Natl. Acad. Sci. USA, 85, 8998-9002 (1988)] is preferably employed. The primers to be used in such PCR method can be appropriately designed based on the sequence information of the gene of the present invention as disclosed herein and can be synthesized by a

-11-

conventional method.

The amplified DNA/RNA fragment can be isolated and purified by a conventional method as mentioned above, for example by gel electrophoresis.

5 The nucleotide sequence of the thus-obtained gene of the present invention or any of various DNA fragments can be determined by a conventional method, for example the dideoxy method [Proc. Natl. Acad. Sci. USA, 74, 5463-5467 (1977)] or the Maxam-Gilbert method
10 [Methods in Enzymology, 65, 499 (1980)]. Such nucleotide sequence determination can be readily performed using a commercially available sequence kit as well.

When the gene of the present invention is used and conventional techniques of recombinant DNA technology
15 [see e.g. Science, 224, p. 1431 (1984); Biochem. Biophys. Res. Comm., 130, p. 692 (1985); Proc. Natl. Acad. Sci. USA, 80, p. 5990 (1983) and the references cited above] are followed, a recombinant protein can be obtained. More detailedly, said protein can be produced by
20 constructing a recombinant DNA enabling the gene of the present invention to be expressed in host cells, introducing it into host cells for transformation thereof and cultivating the resulting transformant.

In that case, the host cells may be eukaryotic
25 or prokaryotic. The eukaryotic cells include vertebrate

-12-

cells, yeast cells and so on, and the vertebrate cells include, but are not limited to, simian cells named COS cells [Cell, 23, 175-182 (1981)], Chinese hamster ovary cells and a dihydrofolate reductase-deficient cell line
5 derived therefrom [Proc. Natl. Acad. Sci. USA, 77, 4216-4220 (1980)] and the like, which are frequently used.

As regards the expression vector to be used with vertebrate cells, an expression vector having a promoter located upstream of the gene to be expressed,
10 RNA splicing sites, a polyadenylation site and a transcription termination sequence can be generally used. This may further have an origin of replication as necessary. As an example of said expression vector, there may be mentioned pSV2dhfr [Mol. Cell. Biol., 1, 854
15 (1981)], which has the SV40 early promoter. As for the eukaryotic microorganisms, yeasts are generally and frequently used and, among them, yeasts of the genus Saccharomyces can be used with advantage. As regards the expression vector for use with said yeasts and other
20 eukaryotic microorganisms, pAM82 [Proc. Natl. Acad. Sci. USA, 80, 1-5 (1983)], which has the acid phosphatase gene promoter, for instance, can be used.

Furthermore, a prokaryotic gene fused vector can be preferably used as the expression vector for the
25 gene of the present invention. As specific examples of

-13-

said vector, there may be mentioned pGEX-2TK and pGEX-4T-2 which have a GST domain (derived from S. japonicum) with a molecular weight of 26,000.

Escherichia coli and Bacillus subtilis are
5 generally and preferably used as prokaryotic hosts. When these are used as hosts in the practice of the present invention, an expression plasmid derived from a plasmid vector capable of replicating in said host organisms and provided in this vector with a promoter and the SD (Shine
10 and Dalgarno) sequence upstream of said gene for enabling the expression of the gene of the present invention and further provided with an initiation codon (e.g. ATG) necessary for the initiation of protein synthesis is preferably used. The Escherichia coli strain K12, among
15 others, is preferably used as the host Escherichia coli, and pBR322 and modified vectors derived therefrom are generally and preferably used as the vector, while various known strains and vectors can also be used. Examples of the promoter which can be used are the
20 tryptophan (trp) promoter, lpp promoter, lac promoter and PL/PR promoter.

The thus-obtained desired recombinant DNA can be introduced into host cells for transformation by using various general methods. The transformant obtained can
25 be cultured by a conventional method and the culture

-14-

leads to expression and production of the desired protein encoded by the gene of the present invention. The medium to be used in said culture can suitably be selected from among various media in conventional use according to the
5 host cells employed. The host cells can be cultured under conditions suited for the growth thereof.

In the above manner, the desired recombinant protein is expressed and produced and accumulated or secreted within the transformant cells or extracellularly
10 or on the cell membrane.

The recombinant protein can be separated and purified as desired by various separation procedures utilizing the physical, chemical and other properties thereof [cf. e.g. "Seikagaku (Biochemistry) Data Book
15 II", pages 1175-1259, 1st Edition, 1st Printing, published June 23, 1980 by Tokyo Kagaku Dojin; Biochemistry, 25 (25), 8274-8277 (1986); Eur. J. Biochem., 163, 313-321 (1987)]. Specifically, said procedures include, among others, ordinary reconstitution treatment,
20 treatment with a protein precipitating agent (salting out), centrifugation, osmotic shock treatment, sonication, ultrafiltration, various liquid chromatography techniques such as molecular sieve chromatography (gel filtration), adsorption chromatography, ion exchange
25 chromatography, affinity chromatography and high-

-15-

performance liquid chromatography (HPLC), dialysis and combinations thereof. Among them, affinity chromatography utilizing a column with the desired protein bound thereto is particularly preferred.

5 Furthermore, on the basis of the sequence information about the gene of the present invention as revealed by the present invention, for example by utilizing part or the whole of said gene, it is possible to detect the expression of the gene of the present
10 invention in various human tissues. This can be performed by a conventional method, for example by RNA amplification by RT-PCR (reverse transcribed-polymerase chain reaction) [Kawasaki, E. S., et al., Amplification of RNA, in PCR Protocol, A guide to methods and
15 applications, Academic Press, Inc., San Diego, 21-27 (1991)], or by northern blotting analysis [Molecular Cloning, Cold Spring Harbor Laboratory (1989)], with good results.

 The primers to be used in employing the above-
20 mentioned PCR method are not limited to any particular ones provided that they are specific to the gene of the present invention and enable the gene of the present invention alone to be specifically amplified. They can be designed or selected appropriately based on the gene
25 information provided by the present invention. They can

-16-

have a partial sequence comprising about 20 to 30 nucleotides according to the established practice. Suitable examples are as shown in Examples 1 to 11.

Thus, the present invention also provides
5 primers and/or probes useful in specifically detecting such novel gene.

By using the novel gene provided by the present invention, it is possible to detect the expression of said gene in various tissues, analyze the structure and
10 function thereof and, further, produce the human protein encoded by said gene in the manner of genetic engineering. These make it possible to analyze the expression product, reveal the pathology of a disease associated therewith, for example a genopathy or cancer,
15 and diagnose and treat the disease.

The following drawings are referred to in the examples.

Fig. 1 shows the result obtained by testing the PI4 kinase activity of NPIK in Example 9. Fig. 2 shows
20 the effect of TritonTM X-100 and adenosine on NPIK activity.

EXAMPLES

The following examples illustrate the present invention in further detail.

25

Example 1

-17-

GDP dissociation stimulator gene

(1) Cloning and DNA sequencing of GDP dissociation stimulator gene

mRNAs extracted from the tissues of human fetal
5 brain, adult blood vessels and placenta were purchased
from Clontech and used as starting materials.

cDNA was synthesized from each mRNA and
inserted into the vector λ ZAPII (Stratagene) to thereby
construct a cDNA library (Otsuka GEN Research Institute,
10 Otsuka Pharmaceutical Co., Ltd.)

Human gene-containing Escherichia coli colonies
were allowed to form on agar medium by the in vivo
excision technique [Short, J. M., et al., Nucleic Acids
Res., 16, 7583-7600 (1988)]. Colonies were picked up at
15 random and human gene-containing Escherichia coli clones
were registered on 96-well micro plates. The clones
registered were stored at -80°C.

Each of the clones registered was cultured
overnight in 1.5 ml of LB medium, and DNA was extracted
20 and purified using a model PI-100 automatic plasmid
extractor (Kurabo). Contaminant Escherichia coli RNA was
decomposed and removed by RNase treatment. The DNA was
dissolved to a final volume of 30 μ l. A 2- μ l portion was
used for roughly checking the DNA size and quantity using
25 a minigel, 7 μ l was used for sequencing reactions and the

-18-

remaining portion (21 μ l) was stored as plasmid DNA at 4°C.

This method, after slight changes in the program, enables extraction of the cosmid, which is
5 useful also as a probe for FISH (fluorescence in situ hybridization) shown later in the examples.

Then, the dideoxy terminator method of Sanger et al. [Sanger, F., et al., Proc. Natl. Acad. Sci. USA, 74, 5463-5467 (1977)] using T3, T7 or a synthetic
10 oligonucleotide primer or the cycle sequence method [Carothers, A. M., et al., Bio. Techniques, 7, 494-499 (1989)] comprising the dideoxy chain terminator method plus PCR method was carried out. These are methods of
terminating the extension reaction specifically to the
15 four bases using a small amount of plasmid DNA (about 0.1 to 0.5 μ g) as a template.

The sequence primers used were FITC (fluorescein isothiocyanate)-labeled ones. Generally, about 25 cycles of reaction were performed using Taq
20 polymerase. The PCR products were separated on a polyacrylamide urea gel and the fluorescence-labeled DNA fragments were submitted to an automatic DNA sequencer (ALFTM DNA Sequencer; Pharmacia) for determining the
sequence of about 400 bases from the 5' terminus side of
25 cDNA.

-19-

Since the 3' nontranslational region is high in heterogeneity for each gene and therefore suited for discriminating individual genes from one another, sequencing was performed on the 3' side as well depending
5 on the situation.

The vast sum of nucleotide sequence information obtained from the DNA sequencer was transferred to a 64-bit DEC 3400 computer for homology analysis by the computer. In the homology analysis, a data base
10 (GenBank, EMBL) was used for searching according to the UWCGG FASTA program [Pearson, W. R. and Lipman, D. J., Proc. Natl. Acad. Sci. USA, 85, 2444-2448 (1988)].

As a result of arbitrary selection by the above method and of cDNA sequence analysis, a clone designated
15 as GEN-501D08 and having a 0.8 kilobase insert was found to show a high level of homology to the C terminal region of the human Ral guanine nucleotide dissociation stimulator (RalGDS) gene. Since RalGDS is considered to play a certain role in signal transmission pathways, the
20 whole nucleotide sequence of the cDNA insert portion providing the human homolog was further determined.

Low-molecular GTPases play an important role in transmitting signals for a number of cell functions including cell proliferation, differentiation and
25 transformation [Bourne, H. R. et al., Nature, 348, 125-

-20-

132 (1990); Bourne et al., *Nature*, 349, 117-127 (1991)].

It is well known that, among them, those proteins encoded by the ras gene family function as molecular switches or, in other words, the functions of the ras gene family are regulated by different conditions of binding proteins such as biologically inactive GDP-binding proteins or active GDP-binding proteins, and that these two conditions are induced by GTPase activating proteins (GAPs) or GDS. The former enzymes induce GDP binding by stimulating the hydrolysis of bound GTP and the latter enzyme induces the regular GTP binding by releasing bound GDP [Boguski, M. S. and McCormick, F., *Nature*, 366, 643-654 (1993)].

RalGDS was first discovered as a member of the ras gene family lacking in transforming activity and as a GDP dissociation stimulator specific to RAS [Chardin, P. and Tavittian, A., *EMBO J.*, 5, 2203-2208 (1986); Albright, C. F., et al., *EMBO J.*, 12, 339-347 (1993)].

In addition to Ral, RalGDS was found to function, through interaction with these proteins, as an effector molecule for N-ras, H-ras, K-ras and Rap [Spaargaren, M. and Bischoff, J. R., *Proc. Natl. Acad. Sci. USA*, 91, 12609-12613 (1994)].

The nucleotide sequence of the cDNA clone designated as GEN-501D08 is shown under SEQ ID NO:3, the

-21-

nucleotide sequence of the coding region of said clone under SEQ ID NO:2, and the amino acid sequence encoded by said nucleotide sequence under SEQ ID NO:1.

5 This cDNA comprises 842 nucleotides, including an open reading frame comprising 366 nucleotides and coding for 122 amino acids. The translation initiation codon was found to be located at the 28th nucleotide residue.

10 Comparison between the RalGDS protein known among conventional databases and the amino acid sequence deduced from said cDNA revealed that the protein encoded by this cDNA is homologous to the C terminal domain of human RalGDS. The amino acid sequence encoded by this novel gene was found to be 39.5% identical with the C
15 terminal domain of RalGDS which is thought to be necessary for binding to ras.

Therefore, it is presumable, as mentioned above, that this gene product might interact with the ras family proteins or have influence on the ras-mediated
20 signal transduction pathways. However, this novel gene is lacking in the region coding for the GDS activity domain and the corresponding protein seems to be different in function from the GDS protein. This gene was named human RalGDS by the present inventors.

25 (2) Northern blot analysis

-22-

The expression of the RalGDS protein mRNA in normal human tissues was evaluated by Northern blotting using, as a probe, the human cDNA clone labeled by the random oligonucleotide priming method.

5 The Northern blot analysis was carried out with a human MTN blot (Human Multiple Tissue Northern blot; Clontech, Palo Alto, CA, USA) according to the manufacturer's protocol.

10 Thus, the PCR amplification product from the above GEN-501D08 clone was labeled with [³²P]-dCTP (random-primed DNA labeling kit, Boehringer-Mannheim) for use as a probe.

15 For blotting, hybridization was performed overnight at 42°C in a solution comprising 50% formamide/5 x SSC/50 x Denhardt's solution/0.1% SDS (containing 100 µg/ml denatured salmon sperm DNA). After washing with two portions of 2 x SSC/0.01% SDS at room temperature, the membrane filter was further washed three times with 0.1 x SSC/0.05% SDS at 50°C for 40 minutes.

20 An X-ray film (Kodak) was exposed to the filter at -70°C for 18 hours.

25 As a result, it was revealed that a 900-bp transcript had been expressed in all the human tissues tested. In addition, a 3.2-kb transcript was observed specifically in the heart and skeletal muscle. The

-23-

expression of these transcripts differing in size may be due either to alternative splicing or to cross hybridization with homologous genes.

(3) Cosmid clone and chromosome localization by FISH

5 FISH was performed by screening a library of human chromosomes cloned in the cosmid vector pWE15 using, as a probe, the 0.8-kb insert of the cDNA clone [Sambrook, J., et al., Molecular Cloning, 2nd Ed., pp. 3.1-3.58, Cold Spring Harbor Laboratory Press, Cold
10 Spring Harbor, New York (1989)].

FISH for chromosome assignment was carried out by the method of Inazawa et al. which comprises G-banding pattern comparison for confirmation [Inazawa, J., et al., Genomics, 17, 153-162 (1993)].

15 For use as a probe, the cosmid DNA (0.5 µg) obtained from chromosome screening and corresponding to GEN-501D08 was labeled with biotin-16-dUTP by nick translation.

To eliminate the background noise due to
20 repetitive sequences, 0.5 µl of sonicated human placenta DNA (10 mg/ml) was added to 9.5 µl of the probe solution. The mixture was denatured at 80°C for 5 minutes and admixed with an equal volume of 4 x SSC containing 20% dextransulfate. Then, a denatured slide was sown with
25 the hybridization mixture and, after covering with

-24-

paraffin, incubated in a wet chamber at 37°C for 16 to 18 hours. After washing with 50% formamide/2 x SSC at 37°C for 15 minutes, the slide was washed with 2 x SSC for 15 minutes and further with 1 x SSC for 15 minutes.

5 The slide was then incubated in 4 x SSC supplemented with "1% Block Ace" (trademark; Dainippon Pharmaceutical) containing avidin-FITC (5 µg/ml) at 37°C for 40 minutes. Then, the slide was washed with 4 x SSC for 10 minutes and with 4 x SSC containing 0.05% Triton X-100
10 for 10 minutes and immersed in an antifading PPD solution [prepared by adjusting 100 mg of PPD (Wako Catalog No. 164-015321) and 10 ml of PBS(-) (pH 7.4) to pH 8.0 with 0.5 M Na₂CO₃/0.5 M NaHCO₃ (9:1, v/v) buffer (pH 9.0) and adding glycerol to make a total volume of 100 ml]
15 containing 1% DABCO [1% DABCO (Sigma) in PBS(-):glycerol 1:9 (v:v)], followed by counter staining with DAPI (4,6-diamino-2-phenylindole; Sigma).

 With more than 100 tested cells in the metaphase, a specific hybridization signal was observed
20 on the chromosome band at 6p21.3, without any signal on other chromosomes. It was thus confirmed that the RalGDS gene is located on the chromosome 6p21.3.

 By using the novel human RalGDS-associated gene of the present invention as obtained in this example, the
25 expression of said gene in various tissues can be

-25-

detected and the human RalGDS protein can be produced in the manner of genetic engineering. These are expected to enable studies on the roles of the expression product protein and ras-mediated signals in transduction pathways as well as pathological investigations of diseases in which these are involved, for example cancer, and the diagnosis and treatment of such diseases. Furthermore, it becomes possible to study the development and progress of diseases involving the same chromosomal translocation of the RalGDS protein gene of the present invention, for example tonic spondylitis, atrial septal defect, pigmentary retinopathy, aphasia and the like.

Example 2

Cytoskeleton-associated protein 2 gene (CKAP2 gene)

(1) Cytoskeleton-associated protein 2 gene cloning and DNA sequencing

cDNA clones were arbitrarily chosen from a human fetal brain cDNA library in the same manner as in Example 1 were subjected to sequence analysis and, as a result, a clone having a base sequence containing the CAP-glycine domain of the human cytoskeleton-associated protein (CAP) gene and highly homologous to several CAP family genes was found and named GEN-080G01.

Meanwhile, the cytoskeleton occurs in the cytoplasm and just inside the cell membrane of eukaryotic

-26-

cells and is a network structure comprising complicatedly entangled filaments. Said cytoskeleton is constituted of microtubules composed of tubulin, microfilaments composed of actin, intermediate filaments composed of desmin and vimentin, and so on. The cytoskeleton not only acts as supportive cellular elements but also isokinetically functions to induce morphological changes of cells by polymerization and depolymerization in the fibrous system. The cytoskeleton binds to intracellular organelles, cell membrane receptors and ion channels and thus plays an important role in intracellular movement and locality maintenance thereof and, in addition, is said to have functions in activity regulation and mutual information transmission. Thus it supposedly occupies a very important position in physiological activity regulation of the whole cell. In particular, the relation between canceration of cells and qualitative changes of the cytoskeleton attracts attention since cancer cells differ in morphology and recognition response from normal cells.

The activity of this cytoskeleton is modulated by a number of cytoskeleton-associated proteins (CAPs). One group of CAPs is characterized by a glycine motif highly conserved and supposedly contributing to association with microtubules [CAP-GLY domain; Riehemann, K.

-27-

and Song, C., Trends Biochem. Sci., 18, 82-83 (1993)].

Among the members of this group of CAPs, there are CLIP-170, 150 kDa DAP (dynein-associated protein, or dynactin), D. melanogaster GLUED, S. cerevisiae BIK1, 5 restin [Bilbe, G., et al., EMBO J., 11, 2103-2113 (1992)]; Hilliker, C., et al., Cytogenet. Cell Genet., 65, 172-176 (1994)] and C. elegans 13.5 kDa protein [Wilson, R., et al., Nature, 368, 32-38 (1994)]. Except for the last two proteins, direct or indirect evidences 10 have suggested that they could interact with microtubules.

The above-mentioned CLIP-170 is essential for the in vitro binding of endocytic vesicles to microtubules and colocalizes with endocytic organelles 15 [Rickard, J. E. and Kreis, T. E., J. Biol. Chem., 18, 82-83 (1990); Pierre, P., et al., Cell, 70, 887-900 (1992)].

The above-mentioned dynactin is one of the factors constituting the cytoplasmic dynein motor, which functions in retrograde vesicle transport [Schroer, T. A. 20 and Sheetz, M. P., J. Cell Biol., 115, 1309-1318 (1991)] or probably in the movement of chromosomes during mitosis [Pfarr, C. M., et al., Nature, 345, 263-265 (1990); Steuer, E. R., et al., Nature, 345, 266-268 (1990); Wordeman, L., et al., J. Cell Biol., 114, 285-294 25 (1991)].

-28-

GLUED, the Drosophila homolog of mammalian dynactin, is essential for the viability of almost all cells and for the proper organization of some neurons [Swaroop, A., et al., Proc. Natl. Acad. Sci. USA, 84, 6501-6505 (1987); Holzbaaur, E. L. P., et al., Nature, 351, 579-583 (1991)].

BIK1 interacts with microtubules and plays an important role in spindle formation during mitosis in yeasts [Trueheart, J., et al., Mol. Cell. Biol., 7, 2316-2326 (1987); Berlin, V., et al., J. Cell Biol., 111, 2573-2586 (1990)].

At present, these genes are classified under the term CAP family (CAPs).

As a result of database searching, the above-mentioned cDNA clone of 463-bp (excluding the poly-A signal) showed significant homology in nucleotide sequence with the restin and CLIP-170 encoding genes. However, said clone was lacking in the 5' region as compared with the restin gene and, therefore, the technique of 5' RACE [Frohman, M. A., et al., Proc. Natl. Acad. Sci. USA, 85, 8998-9002 (1988)] was used to isolate this missing segment.

(2) 5' RACE (5' rapid amplification of cDNA ends)

A cDNA clone containing the 5' portion of the gene of the present invention was isolated for analysis

-29-

by the 5' RACE technique using a commercial kit (5'-Rapid
TM
 AmpliFinder RACE kit, Clontech) according to the
 manufacturer's protocol with minor modifications, as
 follows.

5 The gene-specific primer P1 and primer P2 used
 here were synthesized by the conventional method and
 their nucleotide sequences are as shown below in Table 1.
 The anchor primer used was the one attached to the
 commercial kit.

10

Table 1

Primer	Nucleotide sequence
15 Primer P1	5'-ACACCAATCCAGTAGCCAGGCTTG-3'
Primer P2	5'-CACTCGAGAATCTGTGAGACCTACATACATGACG-3'

cdNA was obtained by reverse transcription of
 0.1 µg of human fetal brain poly(A)+RNA by the random
 20 hexamer technique using reverse transcriptase
 (SuperscriptTM II, Life Technologies) and the cdNA was
 amplified by the first PCR using the P1 primer and anchor
 primer according to Watanabe et al. [Watanabe, T., et
 al., Cell Genet., in press).

25 Thus, to 0.1 µg of the above-mentioned cdNA
 were added 2.5 mM dNTP/1 x Taq buffer (Takara Shuzo)/0.2
 µM P1 primer, 0.2 µM adaptor primer/0.25 unit ExTaqTM
 enzyme (Takara Shuzo) to make a total volume of 50 µl,
 followed by addition of the anchor primer. The mixture

-30-

was subjected to PCR. Thus, 35 cycles of amplification were performed under the conditions: 94°C for 45 seconds, 60°C for 45 seconds, and 72°C for 2 minutes. Finally, the mixture was heated at 72°C for 5 minutes.

5 Then, 1 µl of the 50-µl first PCR product was subjected to amplification by the second PCR using the specific nested P2 primer and anchor primer. The second PCR product was analyzed by 1.5% agarose gel electrophoresis.

10 Upon agarose gel electrophoresis, a single band, about 650 nucleotides in size, was detected. The product from this band was inserted into a vector (pT7Blue(R)T-Vector, Novagen) and a plurality of clones with an insert having an appropriate size were selected.

15 Six of the 5' RACE clones obtained from the PCR product had the same sequence but had different lengths. By sequencing two overlapping cDNA clones, GEN-080G01 and GEN-080G0149, the protein-encoding sequence and 5' and 3' flanking sequences, 1015 nucleotides in total length,
20 were determined. Said gene was named cytoskeleton-associated protein 2 gene (CKAP2 gene).

 The nucleotide sequence obtained from the above-mentioned two overlapping cDNA clones GEN-080G01 and GEN-080G0149 is shown under SEQ ID NO:6, the
25 nucleotide sequence of the coding region of said clone

-31-

under SEQ ID NO:5, and the amino acid sequence encoded by said nucleotide sequence under SEQ ID NO:4.

As shown under SEQ ID NO:6, the CKAP2 gene had a relatively GC-rich 5' noncoding region, with incomplete triplet repeats, (CAG)₄(CGG)₄(CTG)(CGG), occurring at nucleotides 40-69.

ATG located at nucleotides 274-276 is the presumable start codon. A stop codon (TGA) was situated at nucleotides 853-855. A polyadenylation signal (ATTAAA) was followed by 16 nucleotides before the poly(A) start. The estimated open reading frame comprises 579 nucleotides coding for 193 amino acid residues with a calculated molecular weight of 21,800 daltons.

The coding region was further amplified by RT-PCR, to eliminate the possibility of the synthetic sequence obtained being a cDNA chimera.

(2) Similarity of CKAP2 to other CAPs

While sequencing of CKAP2 revealed homology with the sequences of restin and CLIP-170, the homologous region was limited to a short sequence corresponding to the CAP-GLY domain. On the amino acid level, the deduced CKAP2 was highly homologous to five other CAPs in this domain.

CKAP2 was lacking in such other motif

-32-

characteristics of some CAPs as the alpha helical rod and zinc finger motif. The alpha helical rod is thought to contribute to dimerization and to increase the microtubule binding capacity [Pierre, P., et al., Cell, 70, 887-900 (1992)]. The lack of the alpha helical domain might mean that CKAP2 be incapable of homo or hetero dimer formation.

Paralleling of the CAP-GLY domains of these proteins revealed that other conserved residues other than glycine residues are also found in CKAP2. CAPs having a CAP-GLY domain are thought to be associated with the activities of cellular organelles and the interactions thereof with microtubules. Since it contains a CAP-GLY domain, as mentioned above, CKAP2 is placed in the family of CAPs.

Studies with mutants of Glued have revealed that the Glued product plays an important role in almost all cells [Swaroop, A., et al., Proc. Natl. Acad. Sci. USA, 84, 6501-6505 (1987)] and that it has other neuron-specific functions in neuronal cells [Meyerowitz, E. M. and Kankel, D. R., Dev. Biol., 62, 112-142 (1978)]. These microtubule-associated proteins are thought to function in vesicle transport and mitosis. Because of the importance of the vesicle transport system in neuronal cells, defects in these components might lead to

-33-

aberrant neuronal systems.

In view of the above, CKAP2 might be involved in specific neuronal functions as well as in fundamental cellular functions.

5 (3) Northern blot analysis

The expression of human CKAP2 mRNA in normal human tissues was examined by Northern blotting in the same manner as in Example 1 (2) using the GEN-080G01 clone (corresponding to nucleotides 553-1015) as a probe.

10 As a result, in all the eight tissues tested, namely human heart, brain, placenta, lung, liver, skeletal muscle, kidney and pancreas, a 1.0 kb transcript agreeing in size with the CKAP2 cDNA was detected. Said 1.0 kb transcript was expressed at significantly higher
15 levels in heart and brain than in the other tissues examined. Two weak bands, 3.4 kb and 4.6 kb, were also detected in all the tissues examined.

According to the Northern blot analysis, the 3.4 kb and 4.6 kb transcripts might possibly be derived
20 from the same gene coding for the 1.0 kb CKAP2 by alternative splicing or transcribed from other related genes. These characteristics of the transcripts may indicate that CKAP2 might also code for a protein having a CAP-GLY domain as well as an alpha helix.

25 (4) Cosmid cloning and chromosomal localization by

-34-

direct R-banding FISH

Two cosmids corresponding to the CKAP2 cDNA were obtained. These two cosmid clones were subjected to direct R-banding FISH in the same manner as in Example 1
5 (3) for chromosomal locus mapping of CKAP2.

For suppressing the background due to repetitive sequences, a 20-fold excessive amount of human Cot-I DNA (BRL) was added as described by Lichter et al. [Lichter, P., et al., Proc. Natl. Acad. Sci. USA, 87,
10 6634-6638 (1990)]. A Provia 100 film (Fuji ISO 100; Fuji Photo Film) was used for photomicrography.

As a result, CKAP2 was mapped on chromosome bands 19q13.11-q13.12.

Two autosomal dominant neurological diseases
15 have been localized to this region by linkage analysis: CADASIL (cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy) between the DNA markers D19S221 and D19S222, and FHM (familial hemiplegic migraine) between D19S215 and D19S216. These
20 two diseases may be allelic disorders in which the same gene is involved [Tournier-Lasserre, E., et al., Nature Genet., 3, 256-259 (1993); Joutel, A., et al., Nature Genet., 5, 40-45 (1993)].

Although no evidence is available to support
25 CKAP2 as a candidate gene for FHM or CADASIL, it is

By using the novel human CKAP2 gene of the present invention as obtained in this example, it is possible to detect the expression of said gene in various tissues or produce the human CKAP2 gene in the manner of genetic engineering. Through these, it becomes possible to analyze the functions of the human CKAP2 system or human CKAP2, which is involved in diverse activities essential to cells, as mentioned above, to diagnose various neurological diseases in which said system or gene is involved, for example familial migraine, and to screen out and evaluate a therapeutic or prophylactic drug therefor.

OTK27 gene

As a result of sequence analysis of cDNA clones arbitrarily selected from a human fetal brain cDNA library in the same manner as in Example 1 (1) and database searching, a cDNA clone, GEN-025F07, coding for a protein highly homologous to NHP2, a yeast nucleoprotein [Saccharomyces cerevisiae; Kolodrubetz, D. and Burgum, A., YEAST, 7, 79-90 (1991)], was found and named OTK27.

25 Nucleoproteins are fundamental cellular consti-

-36-

tuents of chromosomes, ribosomes and so forth and are thought to play an essential role in cell multiplication and viability. The yeast nucleoprotein NHP2, a high-mobility group (HMG)-like protein, like HMG, has
5 reportedly a function essential for cell viability [Kolodrubetz, D. and Burgum, A., YEAST, 7, 79-90 (1991)].

The novel human gene, OTK27 gene, of the present invention, which is highly homologous to the above-mentioned yeast NHP2 gene, is supposed to be
10 similar in function.

The nucleotide sequence of said GEN-025F07 clone was found to comprise 1493 nucleotides, as shown under SEQ ID NO:9, and contain an open reading frame comprising 384 nucleotides, as shown under SEQ ID NO:8,
15 coding for an amino acid sequence comprising 128 amino acid residues, as shown under SEQ ID NO:7. The initiation codon was located at nucleotides 95-97 of the sequence shown under SEQ ID NO:9, and the termination codon at nucleotides 479-481.

20 At the amino acid level, the OTK27 protein was highly homologous (38%) to NHP2. It was 83% identical with the protein deduced from the cDNA from Arabidopsis thaliana; Newman, T., unpublished; GENEMBL Accession No. T14197).

25 (2) Northern blot analysis

-37-

For examining the expression of human OTK27 mRNA in normal human tissues, the insert in the OTK27 cDNA was amplified by PCR, the PCR product was purified and labeled with [³²P]-dCTP (random-primed DNA labeling kit, Boehringer Mannheim), and Northern blotting was performed using the labeled product as a probe in the same manner as in Example 1 (2).

As a result of the Northern blot analysis, two bands corresponding to possible transcripts from this gene were detected at approximately 1.6 kb and 0.7 kb. Both sizes of transcript were expressed in all normal adult tissues examined. However, the expression of the 0.7 kb transcript was significantly reduced in brain and was of higher levels in heart, skeletal muscle and testicle than in other tissues examined.

For further examination of these two transcripts, eleven cDNA clones were isolated from a testis cDNA library and their DNA sequences were determined in the same manner as in Example 1 (1).

As a result, in six clones, the sequences were found to be in agreement with that of the 0.7 kb transcript, with a poly(A) sequence starting at around the 600th nucleotide, namely at the 598th nucleotide in two of the six clones, at the 606th nucleotide in three clones, and at the 613th nucleotide in one clone.

-38-

In these six clones, the "TATAAA" sequence was recognized at nucleotides 583-588 as a probable poly(A) signal. The upstream poly(A) signal "TATAAA" of this gene was recognized as little influencing in brain and
5 more effective in the three tissues mentioned above than in other tissues. The possibility was considered that the stability of each transcript vary from tissue to tissue.

Results of zoo blot analysis indicated that
10 this gene is well conserved also in other vertebrates. Since this gene is expressed ubiquitously in normal adult tissues and conserved among a wide range of species, the gene product is likely to play an important physiological role. The evidence that yeasts lacking in NHP2 are
15 nonviable suggests that the human homolog may also be essential to cell viability.

(3) Chromosomal localization of OTK27 by direct R-banding FISH

One cosmid clone corresponding to the cDNA
20 OTK27 was isolated from a total human genomic cosmid library (5-genome equivalent) using the OTK27 cDNA insert as a probe and subjected to FISH in the same manner as in Example 1 (3) for chromosomal localization of OTK27.

As a result, two distinct spots were observed
25 on the chromosome band 12q24.3.

-39-

The OTK27 gene of the present invention can be used in causing expression thereof and detecting the OTK27 protein, a human nucleoprotein, and thus can be utilized in the diagnosis and pathologic studies of various diseases in which said protein is involved and, because of its involvement in cell proliferation and differentiation, in screening out and evaluating therapeutic and preventive drugs for cancer.

Example 4

10 OTK18 gene

(1) OTK18 gene cloning and DNA sequencing

Zinc finger proteins are defined as constituting a large family of transcription-regulating proteins in eukaryotes and carry evolutionally conserved structural motifs [Kadonaga, J. T., et al., Cell, 51, 1079-1090 (1987); Klung, A. and Rhodes, D., Trends Biol. Sci., 12, 464-469 (1987); Evans, R. M. and Hollenberg, S. M., Cell, 52, 1-3 (1988)].

The zinc finger, a loop-like motif formed by the interaction between the zinc ion and two residues, cysteine and histidine residues, is involved in the sequence-specific binding of a protein to RNA or DNA. The zinc finger motif was first identified within the amino acid sequence of the Xenopus transcription factor IIIA [Miller, J., et al., EMBO J., 4, 1609-1614 (1986)].

-40-

The C_2H_2 finger motif is in general tandemly repeated and contains an evolutionally conserved intervening sequence of 7 or 8 amino acids. This intervening stretch was first identified in the Kruppel segmentation gene of Drosophila [Rosenberg, U. B., et al., Nature, 319, 336-339 (1986)]. Since then, hundreds of C_2H_2 zinc finger protein-encoding genes have been found in vertebrate genomes.

As a result of sequence analysis of cDNA clones arbitrarily selected from a human fetal brain cDNA library in the same manner as in Example 1 (1) and database searching, several zinc finger structure-containing clones were identified and, further, a clone having a zinc finger structure of the Kruppel type was found.

Since this clone lacked the 5' portion of the transcript, plaque hybridization was performed with a fetal brain cDNA library using, as a probe, an approximately 1.8 kb insert in the cDNA clone, whereby three clones were isolated. The nucleotide sequences of these were determined in the same manner as in Example 1 (1).

Among the three clones, the one having the largest insert spans 3,754 nucleotides including an open reading frame of 2,133 nucleotides coding for 711 amino acids. It was found that said clone contains a novel

-41-

human gene coding for a peptide highly homologous in the zinc finger domain to those encoded by human ZNF41 and the Drosophila Kruppel gene. This gene was named OTK18 gene (derived from the clone GEN-076C09).

5 The nucleotide sequence of the cDNA clone of the OTK18 gene is shown under SEQ ID NO:12, the coding region-containing nucleotide sequence under SEQ ID NO:11, and the predicted amino acid sequence encoded by said OTK18 gene under SEQ ID NO:10.

10 It was found that the amino acid sequence of OTK18 as deduced from SEQ ID NO:12 contains 13 finger motifs on its carboxy side.

(2) Comparison with other zinc finger motif-containing genes

15 Comparison among OTK18, human ZNF41 and the Drosophila Kruppel gene revealed that each finger motif is for the most part conserved in the consensus sequence CXECGKAFXQKSXLX₂HQRXH.

20 Comparison of the consensus sequence of the zinc finger motifs of OTK18 with those of human ZNF41 and the Drosophila Kruppel gene revealed that the Kruppel type motif is well conserved in the OTK18-encoded protein. However, the sequence similarities were limited to zinc finger domains and no significant homologies were
25 found with regard to other regions.

-42-

The zinc finger domain interacts specifically with the target DNA, recognizing an about 5 bp sequence to thereby bind to the DNA helix [Rhodes, D. and Klug, A., Cell, 46, 123-132 (1986)].

5 Based on the idea that, in view of the above, the multiple module (tandem repetitions of zinc finger) can interact with long stretches of DNA, it is presumable that the target DNA of this gene product containing 13 repeated zinc finger units would be a DNA fragment with a
10 length of approximately 65 bp.

(3) Northern blot analysis

Northern blot analysis was performed as described in Example 1 (2) for checking normal human tissues for expression of the human OTK18 mRNA therein by
15 amplifying the insert of the OTK18 cDNA by PCR, purifying the PCR product, labeling the same with [³²P]-dCTP (random-primed DNA labeling kit, Boehringer Mannheim) and using an MTN blot with the labeled product as a probe.

The results of Northern blot analysis revealed
20 that the transcript of OTK18 is approximately 4.3 kb long and is expressed ubiquitously in various normal adult tissues. However, the expression level in the liver and in peripheral blood lymphocytes seemed to be lower than in other organs tested.

25 (4) Cosmid cloning and chromosomal localization by

-43-

direct R-banding FISH

Chromosomal localization of OTK18 was carried out as described in Example 1 (3).

As a result, complete twin spots were identified with 8 samples while 23 samples showed an incomplete signal or twin spots on either or both homologs. All signals appeared at the q13.4 band of chromosome 19. No twin spots were observed on any other chromosomes.

10 The results of FISH thus revealed that this gene is localized on chromosomal band 19q13.4. This region is known to contain many DNA segments that hybridize with oligonucleotides corresponding to zinc finger domains [Hoovers, J. M. N., et al., Genomics, 12,
15 254-263 (1992)]. In addition, at least one other gene coding for a zinc finger domain has been identified in this region [Marine, J.-C., et al., Genomics, 21, 285-286 (1994)].

20 Hence, the chromosome 19q13 is presumably a site of grouping of multiple genes coding for transcription-regulating proteins.

25 When the novel human OTK18 gene provided by this example is used, it becomes possible to detect expression of said gene in various tissues and produce the human OTK18 protein in the manner of genetic

-44-

engineering. Through these, it is possible to analyze the functions of the human transcription regulating protein gene system or human transcription regulating proteins, which are deeply involved in diverse activities fundamental to cells, as mentioned above, to diagnose various diseases with which said gene is associated, for example malformation or cancer resulting from a developmental or differentiation anomaly, and mental or nervous disorder resulting from a developmental anomaly in the nervous system, and further to screen out and evaluate therapeutic or prophylactic drugs for these diseases.

Example 5

Genes encoding human 26S proteasome constituent P42 protein and P27 protein

(1) Cloning and DNA sequencing of genes respectively encoding human 26S proteasome constituent P42 protein and P27 protein

Proteasome, which is a multifunctional protease, is an enzyme occurring widely in eukaryotes from yeasts to humans and decomposing ubiquitin-binding proteins in cells in an energy-dependent manner. Structurally, said proteasome is constituted of 20S proteasome composed of various constituents with a molecular weight of 21 to 31 kilodaltons and a group of

-45-

PA700 regulatory proteins composed of various constituents with a molecular weight of 30 to 112 kilodaltons and showing a sedimentation coefficient of 22S and, as a whole, occurs as a macromolecule with a
5 molecular weight of about 2 million daltons and a sedimentation coefficient of 26S [Rechsteiner, M., et al., J. Biol. Chem., 268, 6065-6068 (1993); Yoshimura, T., et al., J. Struct. Biol., 111, 200-211 (1993); Tanaka, K., et al., New Biologist, 4, 173-187 (1992)].

10 Despite structural and mechanical analyses thereof, the whole picture of proteasome is not yet fully clear. However, according to studies using yeasts and mice in the main, it reportedly has the functions mentioned below and its functions are becoming more and
15 more elucidated.

 The mechanism of energy-dependent proteolysis in cells starts with selection of proteins by ubiquitin binding. It is not 20S proteasome but 26S proteasome that has ubiquitin-conjugated protein decomposing
20 activity which is ATP-dependent [Chu-Ping et al., J. Biol. Chem., 269, 3539-3547 (1994)]. Hence, human 26S proteasome is considered to be useful in elucidating the mechanism of energy-dependent proteolysis.

 Factors involved in the cell cycle regulation
25 are generally short in half-life and in many cases they

-46-

are subject to strict quantitative control. In fact, it has been made clear that the oncogene products Mos, Myc, Fos and so forth can be decomposed by 26S proteasome in an energy- and ubiquitin-dependent manner [Ishida, N., et al., FEBS Lett., 324, 345-348 (1993); Hershko, A. and Ciechanover, A., Annu. Rev. Biochem., 61, 761-807 (1992)] and the importance of proteasome in cell cycle control is being recognized.

Its importance in the immune system has also been pointed out. It is suggested that proteasome is positively involved in class I major histocompatible complex antigen presentation [Michalek, M. T., et al., Nature, 363, 552-554 (1993)] and it is further suggested that proteasome may be involved in Alzheimer disease, since the phenomena of abnormal accumulation of ubiquitin-conjugated proteins in the brain of patients with Alzheimer disease [Kitaguchi, N., et al., Nature, 361, 530-532 (1988)]. Because of its diverse functions such as those mentioned above, proteasome attracts attention from the viewpoint of its utility in the diagnosis and treatment of various diseases.

A main function of 26S proteasome is ubiquitin-conjugated protein decomposing activity. In particular, it is known that cell cycle-related gene products such as oncogene products and cyclins, typically c-Myc, are

-47-

degraded via ubiquitin-dependent pathways. It has also been observed that the proteasome gene is expressed abnormally in liver cancer cells, renal cancer cells, leukemia cells and the like as compared with normal cells
5 [Kanayama, H., et al., Cancer Res., 51, 6677-6685 (1991)] and that proteasome is abnormally accumulated in tumor cell nuclei. Hence, constituents of proteasome are expected to be useful in studying the mechanism of such canceration and in the diagnosis or treatment of cancer.

10 Also, it is known that the expression of proteasome is induced by interferon γ and so on and is deeply involved in antigen presentation in cells [Aki, M., et al., J. Biochem., 115, 257-269 (1994)]. Hence, constituents of human proteasome are expected to be
15 useful in studying the mechanism of antigen presentation in the immune system and in developing immunoregulating drugs.

Furthermore, proteasome is considered to be deeply associated with ubiquitin abnormally accumulated
20 in the brain of patients with Alzheimer disease. Hence, it is suggested that constituents of human proteasome should be useful in studying the cause of Alzheimer disease and in the treatment of said disease.

In addition to the utilization of expectedly
25 multifunctional proteasome as such in the above manner,

-48-

it is probably possible to produce antibodies using constituents of proteasome as antigens and use such antibodies in diagnosing various diseases by immunoassay. Its utility in this field of diagnosis is thus also a focus of interest.

Meanwhile, a protein having the characteristics of human 26S proteasome is disclosed, for example in Japanese Unexamined Patent Publication No. 292964/1993 and rat proteasome constituents are disclosed in Japanese Unexamined Patent Publication Nos. 268957/1993 and 317059/1993. However, no human 26S proteasome constituents are known. Therefore, the present inventors made a further search for human 26S proteasome constituents and successfully obtained two novel human 26S proteasome constituents, namely human 26S proteasome constituent P42 protein and human S26 proteasome constituent P27 protein, and performed cloning and DNA sequencing of the corresponding genes in the following manner.

(1) Purification of human 26S proteasome constituents
P42 protein and P27 protein

Human proteasome was purified using about 100 g of fresh human kidney and following the method of purifying human proteasome as described in Japanese Unexamined Patent Publication No. 292964/1993, namely by column

-49-

chromatography using BioGel A-1.5 m (5 x 90 cm, Bio-Rad),
hydroxyapatite (1.5 x 15 cm, Bio-Rad) and Q-SepharoseTM
(1.5 x 15 cm, Pharmacia) and glycerol density gradient
centrifugation.

5 The thus-obtained human proteasome was
subjected to reversed phase high performance liquid
chromatography (HPLC) using a Hitachi model L6200 HPLC
system. A Shodex RS Pak D4-613 (0.6 x 15 cm, ShowaTM
Denko) was used and gradient elution was performed with
10 the following two solutions:

First solution: 0.06% trifluoroacetic acid;

Second solution: 0.05% trifluoroacetic acid, 70%
acetonitrile.

 An aliquot of each eluate fraction was
15 subjected to 8.5% SDS-polyacrylamide electrophoresis
under conditions of reduction with dithiothreitol. The
P42 protein and P27 protein thus detected were isolated
and purified.

 The purified P42 and P27 proteins were respec-
20 tively digested with 1 µg of trypsin in 0.1 M Tris buffer
(pH 7.8) containing 2 M urea at 37°C for 8 hours and the
partial peptide fragments obtained were separated by
reversed phase HPLC and their sequences were determined
by Edman degradation. The results obtained are as shown
25 below in Table 2.

-50-

Table 2

Partial protein	Amino acid sequence
P42 (1)	VLNISLW
(2)	TLMELLNQMDGFDLHR
(3)	AVSDFVVSEYXMXA
(4)	EVDPLVYNX
(5)	HGEIDYEAIVK
(6)	LSXGFNGADLRNVXTEAGMFAIXAD
(7)	MIMATNRPDTLDPALLRPGXL
(8)	IHIDLPNEQARLDILK
(9)	ATNGPRYVVVG
(10)	EIDGRLK
(11)	ALQSVGQIVGEVLK
(12)	ILAGPITK
(13)	XXVIELPLTNPELFQG
(14)	VVSSSLVDK
(15)	ALQDYRK
(16)	EHREQLK
(17)	KLESKLDYKPVR
P27 (1)	LVPTR
(2)	AKEEEIEAQIK
(3)	ANYEVLESQK
(4)	VEDALHQLHAR
(5)	DVDLYQVR
(6)	QSQGLSPAQAFK
(7)	AGSQSGGSPEASGVTVDVQE
(8)	GLLGXNI IPLQR

-51-

(2) cDNA library screening, clone isolation and cDNA
nucleotide sequence determination

As mentioned in Example 1 (1), the present
inventors have a database comprising about 30,000 cDNA
5 data as constructed based on large-scale DNA sequencing
using human fetal brain, arterial blood vessel and
placenta cDNA libraries.

Based on the amino acid sequences obtained as
mentioned above in (1), computer searching was performed
10 with the FASTA program (search for homology between said
amino acid sequences and the amino acid sequences
estimated from the database). As regards P42, a clone
(GEN-331G07) showing identity with regard to two amino
acid sequences [(2) and (7) shown in table 2] was
15 screened out and, as regards P27, a clone (GEN-163D09)
showing identity with regard to two amino acid sequences
[(1) and (8) shown in Table 2] was found.

For each of these clones, the 5' side sequence
was determined by 5' RACE and the whole sequence was
20 determined, in the same manner as in Example 2 (2).

As a result, it was revealed that the above-
mentioned P42 clone GEN-331G07 comprises a 1,566-
nucleotide sequence as shown under SEQ ID NO:15,
inclusive of a 1,167-nucleotide open reading frame as
25 shown under SEQ ID NO:14, and that the amino acid

-52-

sequence encoded thereby is the one shown under SEQ ID NO:13 and comprises 389 amino acid residues.

The results of computer homology search revealed that the P42 protein is significantly homologous to the AAA (ATPase associated with a variety of cellular activities) protein family (e.g. P45, TBP1, TBP7, S4, MSS1, etc.). It was thus suggested that it is a new member of the AAA protein family.

As for the P27 clone GEN-163D09, it was revealed that it comprises a 1,128-nucleotide sequence as shown under SEQ ID NO:18, including a 669-nucleotide open reading frame as shown under SEQ ID NO:17 and that the amino acid sequence encoded thereby is the one shown under SEQ ID NO:16 and comprises 223 amino acid residues.

As regards the P27 protein, homology search using a computer failed to reveal any homologous gene among public databases. Thus, the gene in question is presumably a novel gene having an unknown function.

Originally, the above-mentioned P42 and P27 gene products were both purified as regulatory subunit components of proteasome complex. Therefore, these are expected to play an important role in various biological functions through proteolysis, for example a role in energy supply through decomposition of ATP and, hence, they are presumably useful not only in studying the

-53-

function of human 26S proteasome but also in the diagnosis and treatment of various diseases caused by lowering of said biological functions, among others.

Example 6

5 BNAP gene

(1) BNAP gene cloning and DNA sequencing

The nucleosome composed of DNA and histone is a fundamental structure constituting chromosomes in eukaryotic cells and is well conserved over borders among
10 species. This structure is closely associated with the processes of replication and transcription of DNA. However, the nucleosome formation is not fully understood as yet. Only certain specific factors involved in nucleosome assembly (NAPs) have been identified. Thus,
15 two acidic proteins, nucleoplasmin and N1, are already known to facilitate nucleosome construction [Kleinschmidt, J. A., et al., J. Biol. Chem., 260, 1166-1176 (1985); Dilworth, S. M., et al., Cell, 51, 1009-1018 (1987)].

20 A yeast gene, NAP-I, was isolated using a monoclonal antibody and recombinant proteins derived therefrom were tested as to whether they have nucleosome assembling activity in vivo.

More recently, a mouse NAP-I gene, which is a
25 mammalian homolog of the yeast NAP-I gene was cloned

-54-

(Okuda, A.; registered in database under the accession number D12618). Also cloned were a mouse gene, DN38 [Kato, K., Eur. J. Neurosci., 2, 704-711 (1990)] and a human nucleosome assembly protein (hNRP) [Simon, H. U., et al., Biochem. J., 297, 389-397 (1994)]. It was shown that the hNRP gene is expressed in many tissues and is associated with T lymphocyte proliferation.

The present inventors performed sequence analysis of cDNA clones arbitrarily chosen from a human fetal brain cDNA library in the same manner as in Example 1 (1), followed by searches among databases and, as a result, made it clear that a 1,125-nucleotide cDNA clone (free of poly(A)), GEN-078D05, is significantly homologous to the mouse NAP-I gene, which is a gene for a nucleosome assembly protein (NAP) involved in nucleosome construction, a mouse partial cDNA clone, DN38, and hNRP.

Since said clone GEN-078D05 was lacking in the 5' region, 5' RACE was performed in the same manner as in Example 2 (2) to obtain the whole coding region. For this 5' RACE, primers P1 and P2 respectively having the nucleotide sequences shown below in Table 3.

Table 3

Primer	Nucleotide sequence
Primer P1	5'-TTGAAGAATGATGCATTAGGAACCAC-3'
Primer P2	5'-CACTCGAGTGGCTGGATTTC AATTTCTCCAGTAG-3'

-55-

After the first 5' RACE, a single band corresponding to a sequence length of 1,300 nucleotides was obtained. This product was inserted into pT7Blue(R) T-Vector and several clones appropriate in insert size
5 were selected.

Ten 5' RACE clones obtained from two independent PCR reactions were sequenced and the longest clone GEN-078D05TA13 (about 1,300 nucleotides long) was further analyzed.

10 Both strands of the two overlapping cDNA clones GEN-078D05 and GEN-078D05TA13 were sequenced, whereby it was confirmed that the two clones did not yet cover the whole coding region. Therefore, a further second 5' RACE was carried out. For the second 5' RACE, two primers, P3
15 and P4, respectively having the sequences shown below in Table 4 were used.

Table 4

20	Primer	Nucleotide sequence
	Primer P3	5'-GTCGAGCTAGCCATCTCCTCTTCG-3'
	Primer P4	5'-CATGGGCGACAGGTTCCGAGACC-3'

25 A clone, GEN-078D0508, obtained by the second 5' RACE was 300 nucleotides long. This clone contained an estimable initiation codon and three preceding in-frame termination codons. From these three overlapping clones, it became clear that the whole coding region

-56-

comprises 2,636 nucleotides. This gene was named brain-specific nucleosome assembly protein (BNAP) gene.

The BNAP gene contains a 1,518-nucleotide open reading frame shown under SEQ ID NO:20. The amino acid
5 encoded thereby comprises 506 amino acid residues, as shown under SEQ ID NO:19, and the nucleotide sequence of the whole cDNA clone of BNAP is as shown under SEQ ID NO:21.

As shown under SEQ ID NO:21, the 5' noncoding
10 region of said gene was found to be generally rich in GC. Candidate initiation codon sequences were found at nucleotides Nos. 266-268, 287-289 and 329-331. These three sequences all had well conserved sequences in the vicinity of the initiation codons [Kozak, M., J. Biol.
15 Chem., 266, 19867-19870 (1991)].

According to the scanning model, the first ATG (nucleotides Nos. 266-268) of the cDNA clone may be the initiation codon. The termination codon was located at nucleotides Nos. 1784-1786.

20 The 3' noncoding region was generally rich in AT and two polyadenylation signals (AATAAA) were located at nucleotides Nos. 2606-2611 and 2610-2615, respectively.

The longest open reading frame comprised 1,518
25 nucleotides coding for 506 amino acid residues and the

-57-

calculated molecular weight of the BNAP gene product was 57,600 daltons.

Hydrophilic plots indicated that BNAP is very hydrophilic, like other NAPs.

5 For recombinant BNAP expression and purification and for eliminating the possibility that the BNAP gene sequence might give three chimera clones in the step of 5' RACE, RT-PCR was performed using a sequence comprising nucleotides Nos. 326-356 as a sense primer and
10 a sequence comprising nucleotides Nos. 1758-1786 as an antisenses primer.

As a result, a single product of about 1,500 bp was obtained and it was thus confirmed that said sequence is not a chimera but a single transcript.

15 (2) Comparison between BNAP and NAPs

The amino acid sequence deduced from BNAP showed 46% identity and 65% similarity to hNRP.

The deduced BNAP gene product had motifs characteristic of the NAPs already reported and of BNAP.
20 In general, half of the C terminus was well conserved in humans and yeasts.

The first motif (domain I) is KGIPDYWLI (corresponding to amino acid residues Nos. 309-317). This was observed also in hNRP (KGIPSFWLT) and in yeast NAP-I
25 (KGIPSEFWLT).

-58-

The second motif (domain II) is ASFFNFFSPP (corresponding to amino acid residues Nos. 437-446) and this was expressed as DSFFNFFAPP in hNRP and as ESFFNFFSP in yeast NAP-I.

5 These two motifs were also conserved in the deduced mouse NAP-I and DN38 peptides. Both conserved motifs were each a hydrophilic cluster, and the Cys in position 402 was also found conserved.

10 Half of the N terminus had no motifs strictly conserved from yeasts to mammalian species, while motifs conserved among mammalian species were found.

15 For instance, HDLERKYA (corresponding to amino acid residues Nos. 130 to 137) and IINAEYEPTEEEECEW (corresponding to amino acid residues Nos. 150-164), which may be associated with mammal-specific functions, were found strictly conserved.

20 NAPs had acidic stretches, which are believed to be readily capable of binding to histone or other basic proteins. All NAPs had three acidic stretches but the locations thereof were not conserved.

25 BNAP has no such three acidic stretches but, instead, three repeated sequences (corresponding to amino acid residues Nos. 194-207, 208-221 and 222-235) with a long acidic cluster, inclusive of 41 amino acid residues out of 98 amino acid residues, the consensus sequence

-59-

being ExxKExPEVKxEEK (each x being a nonconserved, mostly hydrophobic, residue).

Furthermore, it was revealed that the BNAP sequence had several BNAP-specific motifs. Thus, an
5 extremely serine-rich domain (corresponding to amino acid residues Nos. 24-72) with 33 (67%) of 49 amino acid residues being serine residues was found in the N-terminus portion. On the nucleic acid level, they were reflected as incomplete repetitions of AGC.

10 Following this serine-rich region, there appeared a basic domain (corresponding to amino acid residues Nos. 71-89) comprising 10 basic amino acid residues among 19 residues.

BNAP is supposed to be localized in the
15 nucleus. Two possible signals localized in the nucleus were observed (NLSs). The first signal was found in the basic domain of BNAP and its sequence YRKRR (corresponding to amino acid residues Nos. 75-79) was similar to NLS (GRKKR) of Tat of HIV-1. The second signal was
20 located in the C terminus and its sequence KKYRK (corresponding to amino acid residues Nos. 502-506) was similar to NLS (KKKRR) of the large T antigen of SV40. The presence of these two presumable NLSs suggested the localization of BNAP in the nucleus. However the
25 possibility that other basic clusters might act as NLSs

-60-

could not be excluded.

BNAP has several phosphorylation sites and the activity of BNAP may be controlled through phosphorylation thereof.

5 (3) Northern blot analysis

Northern blot analysis was performed as described in Example 1 (2). Thus, the clone GEN-078D05TA13 (corresponding to nucleotides Nos. 323 to 1558 in the BNAP gene sequence) was amplified by PCR, the PCR
10 product was purified and labeled with [³²P]-dCTP (random-primed DNA labeling kit, Boehringer Mannheim), and the expression of BNAP mRNA in normal human tissues was examined using an MTN blot with the labeled product as a probe.

15 As a result of Northern blot analysis, a 3.0 kb transcript of BNAP was detected (8-hour exposure) in the brain among eight human adult tissues tested, namely heart, brain, placenta, lung, liver, skeletal muscle, kidney and pancreas and, after longer exposure (24
20 hours), a dim band of the same size was detected in the heart.

BNAP was found equally expressed in several sites of brain tested whereas, in other tissues, no signal was detected at all even after 72 hours of
25 exposure. hNRP mRNA was found expressed everywhere in

-61-

the human tissues tested whereas the expression of BNAP mRNA was tissue-specific.

(4) Radiation hybrid mapping

Chromosomal mapping of the BNAP clone was performed by means of radiation hybrid mapping [Cox, D. R., et al., Science, 250, 245-250 (1990)].

Thus, a total human genome radiation hybrid clone (G3RH) panel was purchased from Research Genetics, Inc., AL, USA and PCR was carried out for chromosomal mapping analysis according to the product manual using two primers, A1 and A2, respectively having the nucleotide sequences shown in Table 5.

Table 5

15	Primer	Nucleotide sequence
	A1 primer	5'-CCTAAAAAGTGTCTAAGTGCCAGTT-3'
20	A2 primer	5'-TCAGTGAAAGGGAAGGTAGAACAC-3'

The results obtained were analyzed utilizing softwares usable on the Internet [Boehnke, M., et al., Am. J. Hum. Genet., 46, 581-586 (1991)].

As a result, the BNAP gene was found strongly linked to the marker DXS990 (LOD = 1000, cR8000 = -0.00). Since DXS990 is a marker localized on the chromosome Xq21.3-q22, it was established that BNAP is localized to the chromosomal locus Xq21.3-q22 where genes involved in several signs or symptoms of X-chromosome-associated

-62-

mental retardation are localized.

The nucleosome is not only a fundamental chromosomal structural unit characteristic of eukaryotes but also a gene expression regulating unit. Several results indicate that genes with high transcription activity are sensitive to nuclease treatment, suggesting that the chromosome structure changes with the transcription activity [Elgin, S. C. R., J. Biol. Chem., 263, 19259-19262 (1988)].

NAP-I has been cloned in yeast, mouse and human and is one of the factors capable of promoting nucleosome construction in vivo. In a study performed on their sequences, NAPs containing the epitope of the specific antibody 4A8 were detected in human, mouse, frog, Drosophila and yeast (Saccharomyces cerevisiae) [Ishimi, Y., et al., Eur. J. Biochem., 162, 19-24 (1987)].

In these experiments, NAPs, upon SDS-PAGE analysis, electrophoretically migrated to positions corresponding to a molecular weight between 50 and 60 kDa, whereas the recombinant BNAP slowly migrated to a position of about 80 kDa. The epitope of 4A8 was shown to be localized in the second, well-conserved, hydrophobic motif. And, it was simultaneously shown that the triplet FNF is important as a part of the epitope [Fujii-Nakata, T., et al., J. Biol. Chem., 267, 20980-

-63-

20986 (1992)].

BNAP also contained this consensus motif in domain II. The fact that domain II is markedly hydrophobic and the fact that domain II can be recognized by the immune system suggest that it is probably presented on the BNAP surface and is possibly involved in protein-protein interactions.

Domain I, too, may be involved in protein-protein interactions. Considering that these are conserved generally among NAPs, though to a relatively low extent, it is conceivable that they must be essential for nucleosome construction, although the functional meaning of the conserved domains is still unknown.

The hNRP gene is expressed in thyroid gland, stomach, kidney, intestine, leukemia, lung cancer, mammary cancer and so on [Simon, H. U., et al., Biochem. J., 297, 389-397 (1994)]. Like that, NAPs are expressed everywhere and are thought to be playing an important role in fundamental nucleosome formation.

BNAP may be involved in brain-specific nucleosome formation and an insufficiency thereof may cause neurological diseases or mental retardation as a result of deviated functions of neurons.

BNAP was found strongly linked to a marker on the X-chromosome q21.3-q22 where sequences involved in

-64-

several symptoms of X-chromosome-associated mental retardation are localized. This center-surrounding region of X-chromosome was rich in genes responsible for α -thalassemia, mental retardation (ATR-X) or some other forms of mental retardation [Gibbons, R. J., et al., Cell, 80, 837-845 (1995)]. Like the analysis of the ATR-X gene which seems to regulate the nucleosome structure, the present inventors suppose that BNAP may be involved in a certain type of X-chromosome-linked mental retardation.

According to this example, the novel BNAP gene is provided and, when said gene is used, it is possible to detect the expression of said gene in various tissues and to produce the BNAP protein by the technology of genetic engineering. Through these, it is possible to study the brain nucleosome formation deeply involved, as mentioned above, in variegated activities essential to cells as well as the functions of cranial nerve cells and to diagnose various neurological diseases or mental retardation in which these are involved and screen out and evaluate drugs for the treatment or prevention of such diseases.

Example 7

Human skeletal muscle-specific ubiquitin-conjugating enzyme gene (UBE2G gene)

-65-

The ubiquitin system is a group of enzymes essential for cellular processes and is conserved from yeast to human. Said system is composed of ubiquitin-activating enzymes (UBAs), ubiquitin-conjugating enzymes (UBCs), ubiquitin protein ligases (UBRs) and 26S proteasome particles.

Ubiquitin is transferred from the above-mentioned UBAs to several UBCs, whereby it is activated. UBCs transfer ubiquitins to target proteins with or without the participation of UBRs. These ubiquitin-conjugated target proteins are said to induce a number of cellular responses, such as protein degradation, protein modification, protein translocation, DNA repair, cell cycle control, transcription control, stress responses, etc. and immunological responses [Jentsch, S., et al., Biochim. Biophys. Acta, 1089, 127-139 (1991); Hershko, A. and Ciechanover, A., Annu. Rev. Biochem., 61, 761-807 (1992); Jentsch, S., Annu. Rev. Genet., 26, 179-207 (1992); Ciechanover, A., Cell, 79, 13-21 (1994)].

UBCs are key components of this system and seem to have distinct substrate specificities and modulate different functions. For example, Saccharomyces cerevisiae UBC7 is induced by cadmium and involved in resistance to cadmium poisoning [Jungmann, J., et al., Nature, 361, 369-371 (1993)]. Degradation of MAT- α 2 is

-66-

also executed by UBC7 and UBC6 [Chen, P., et al., Cell, 74, 357-369 (1993)].

The novel gene obtained in this example is UBC7-like gene strongly expressed in human skeletal muscle. In the following, cloning and and DNA sequencing thereof are described.

(1) Cloning and DNA sequencing of human skeletal muscle-specific ubiquitin-conjugating enzyme gene (UBE2G gene)

Following the same procedure as in Example 1 (1), cDNA clones were arbitrarily selected from a human fetal brain cDNA library and subjected to sequence analysis, and database searches were performed. As a result, a cDNA clone, GEN-423A12, was found to have a significantly high level of homology to the genes coding for ubiquitin-conjugating enzymes (UBCs) in various species.

Since said GEN-423A12 clone was lacking in the 5' side, 5' RACE was performed in the same manner as in Example 2 (2) to obtain an entire coding region.

For said 5' RACE, two primers, P1 and P2, respectively having the nucleotide sequences shown in Table 6 were used.

-67-

Table 6

	Primer	Nucleotide sequence
5	P1 primer	5'-TAATGAATTTTCATTTTAGGAGGTCGG-3'
	P2 primer	5'-ATCTTTTGGGAAAGTAAGATGAGCC-3'

The 5' RACE product was inserted into
 10 pT7Blue(R) T-Vector and clones with an insert proper in
 size were selected.

Four of the 5' RACE clones obtained from two
 independent PCR reactions contained the same sequence but
 were different in length.

15 By sequencing the above clones, the coding
 sequence and adjacent 5'- and 3'-flanking sequences of
 the novel gene were determined.

As a result, it was revealed that the novel
 gene has a total length of 617 nucleotides. This gene
 20 was named human skeletal muscle-specific ubiquitin-
 conjugating enzyme gene (UBE2G gene).

To exclude the conceivable possibility that
 this sequence was a chimera clone, RT-PCR was performed
 in the same manner as in Example 6 (1) using the sense
 25 primer to amplify said sequence from the human fetal
 brain cDNA library. As a result, a single PCR product
 was obtained, whereby it was confirmed that said sequence
 is not a chimera one.

The UBE2G gene contains an open reading frame

-68-

of 510 nucleotides, which is shown under SEQ ID NO:23,
the amino acid sequence encoded thereby comprises 170
amino acid residues, as shown under SEQ ID NO:22, and the
nucleotide sequence of the entire UBE2G cDNA is as shown
5 under SEQ ID NO:24.

As shown under SEQ ID NO:24, the estimable
initiation codon was located at nucleotides Nos. 19-21,
corresponding to the first ATG triplet of the cDNA clone.
Since no preceding in-frame termination codon was found,
10 it was deduced that this clone contains the entire open
reading frame on the following grounds.

Thus, (a) the amino acid sequence is highly
homologous to S. cerevisiae UBC7 and said initiation
codon agrees with that of yeast UBC7, supporting said ATG
15 as such. (b) The sequence AGGATGA is similar to the
consensus sequence (A/G)CCATGG around the initiation
codon [Kozak, M., J. Biol. Chem., 266, 19867-19870
(1991)].

(2) Comparison in amino acid sequence between UBE2G and
20 UBCs

Comparison in amino acid sequence between UBE2G
and UBCs suggested that the active site cystein capable
of binding to ubiquitin should be the 90th residue
cystein. The peptides encoded by these genes seem to
25 belong to the same family.

-69-

(3) Northern blot analysis

Northern blot analysis was carried out as described in Example 1 (2). Thus, the entire sequence of UBE2G was amplified by PCR, the PCR product was purified
5 and labeled with [³²P]-dCTP (random-primed DNA labeling kit, Boehringer Mannheim) and the expression of UBE2G mRNA in normal human tissues using the labeled product as a probe. The membrane used was an MTN blot.

As a result of the Northern blot analysis, 4.4
10 kb, 2.4 kb and 1.6 kb transcripts could be detected in all 16 human adult tissues, namely heart, brain, placenta, lung, liver, skeletal muscle, kidney, pancreas, spleen, thyroid gland, urinary bladder, testis, ovary, small intestine, large intestine and peripheral blood
15 leukocyte, after 18 hours of exposure. Strong expression of these transcripts was observed in skeletal muscle.

(4) Radiation hybrid mapping

Chromosomal mapping of the UBE2G clone was performed by radiation hybrid mapping in the same manner as
20 in Example 6 (4).

The primers C1 and C4 used in PCR for chromosomal mapping analysis respectively correspond to nucleotides Nos. 415-435 and nucleotides Nos. 509-528 in the sequence shown under SEQ ID NO:24 and their
25 nucleotide sequences are as shown below in Table 7.

-70-

Table 7

	Primer	Nucleotide sequence
5	C1 primer	5'-GGAGACTCACCTGCTAATGTT-3'
	C4 primer	5'-CTCAAAAGCAGTCTCTTGGC-3'

As a result, the UBE2G gene was found linked to the markers D1S446 (LOD = 12.52, cR8000 = 8.60) and D1S235 (LOD = 9.14, cR8000 = 22.46). These markers are localized to the chromosome bands 1q42.13-q42.3.

UBE2G was expressed strongly in skeletal muscle and very weakly in all other tissues examined. All other UBCs are involved in essential cellular functions, such as cell cycle control, and those UBCs are expressed ubiquitously. However, the expression pattern of UBE2G might suggest a muscle-specific role thereof.

While the three transcripts differing in size were detected, attempts failed to identify which corresponds to the cDNA clone. The primary structure of the UBE2G product showed an extreme homology to yeast UBC7. On the other hand, nematode UBC7 showed strong homology to yeast UBC7. It is involved in degradation of the repressor and further confers resistance to cadmium in yeasts. The similarities among these proteins suggest that they belong to the same family.

It is speculated that UBE2G is involved in degradation of muscle-specific proteins and that a defect

-71-

in said gene could lead to such diseases as muscular dystrophy. Recently, another proteolytic enzyme, calpain 3, was found to be responsible for limb-girdle muscular dystrophy type 2A [Richard, I., et al., Cell, 81, 27-40 (1995)]. At the present, the chromosomal location of UBE2G suggests no significant relationship with any hereditary muscular disease but it is likely that a relation to the gene will be unearthed by linkage analysis in future.

10 In accordance with this example, the novel UBE2G gene is provided and the use of said gene enables detection of its expression in various tissues and production of the UBE2G protein by the technology of genetic engineering. Through these, it becomes possible
15 to study the degradation of muscle-specific proteins deeply involved in basic activities variegated and essential to cells, as mentioned above, and the functions of skeletal muscle, to diagnose various muscular diseases in which these are involved and further to screen out and
20 evaluate drugs for the treatment and prevention of such diseases.

Example 8

TMP-2 gene

(1) TMP-2 gene cloning and DNA sequencing

25 Following the procedure of Example 1 (1), cDNA

-72-

clones were arbitrarily selected from a human fetal brain cDNA library and subjected to sequence analysis, and database searches were performed. As a result, a clone (GEN-092E10) having a cDNA sequence highly homologous to
5 a transmembrane protein gene (accession No.: U19878) was found out.

Membrane protein genes have so far been cloned in frog (Xenopus laevis) and human. These are considered to be a gene for a transmembrane type protein having a
10 follistatin module and an epidermal growth factor (EGF) domain (accession No.: U19878).

The sequence information of the above protein gene indicated that the GEN-092E10 clone was lacking in the 5' region, so that the λ gt10 cDNA library (human
15 fetal brain 5'-STRETCH PLUS cDNA; Clontech) was screened using the GEN-092E10 clone as a probe, whereby a cDNA clone containing a further 5' upstream region was isolated.

Both strands of this cDNA clone were sequenced,
20 whereby the sequence covering the entire coding region became clear. This gene was named TMP-2 gene.

The TMP-2 gene was found to contain an open reading frame of 1,122 nucleotides, as shown under SEQ ID NO:26, encoding an amino acid sequence of 374 residues,
25 as shown under SEQ ID NO:25. The nucleotide sequence of

-73-

the entire TMP-2 cDNA clone comprises 1,721 nucleotides,
as shown under SEQ ID NO:27.

As shown under SEQ ID NO:27, the 5' noncoding
region was generally rich in GC. Several candidates for
5 the initiation codon were found but, according to the
scanning model, the 5th ATG of the cDNA clone (bases Nos.
368-370) was estimated as the initiation codon. The
termination codon was located at nucleotides Nos. 1490-
1492. The polyadenylation signal (AATAAA) was located at
10 nucleotides Nos. 1703-1708. The calculated molecular
weight of the TMP-2 gene product was 41,400 daltons.

As mentioned above, the transmembrane genes
have a follistatin module and an EGF domain. These
motifs were also found conserved in the novel human gene
15 of the present invention.

The TMP-2 gene of the present invention
presumably plays an important role in cell proliferation
or intercellular communication, since, on the amino acid
level, said gene shows homology, across the EGF domain,
20 to TGF- α (transforming growth factor- α ; Derynck, R., et
al., Cell, 38, 287-297 (1984)), beta-cellulin [Igarashi,
K. and Folkman, J., Science, 259, 1604-1607 (1993)],
heparin-binding EGF-like growth factor [Higashiyama, S.,
et al., Science, 251, 936-939 (1991)] and schwannoma-
25 derived growth factor [Kimura, H., et al., Nature, 348,

-74-

257-260 (1990)].

(2) Northern blot analysis

Northern blot analysis was carried out as described in Example 1 (2). Thus, the clone GEN-092E10 was amplified by PCR, the PCR product was purified and labeled with [³²P]-dCTP (random-primed DNA labeling kit, Boehringer Mannheim), and the expression of TMP-2 mRNA in normal human tissues was examined using an MTN blot with the labeled product as a probe.

As a result, high levels of expression were detected in brain and prostate gland. Said TMP-2 gene mRNA was about 2 kb in size.

According to the present invention, the novel human TMP-2 gene is provided and the use of said gene makes it possible to detect the expression of said gene in various tissues or produce the human TMP-2 protein by the technology of genetic engineering and, through these, it becomes possible to study brain tumor and prostatic cancer, which are closely associated with cell proliferation or intercellular communication, as mentioned above, to diagnose these diseases and to screen out and evaluate drugs for the treatment and prevention of such diseases.

Example 9

Human NPIK gene

-75-

(1) Human NPIK gene cloning and DNA sequencing

Following the procedures of Example 1 and Example 2, cDNA clones were arbitrarily selected from a human fetal brain cDNA library and subjected to sequence analysis, and database searches were performed. As a result, two cDNA clones highly homologous to the gene coding for an amino acid sequence conserved in phosphatidylinositol 3 and 4 kinases [Kunz, J., et al., Cell, 73, 585-596 (1993)] were obtained. These were named GEN-428B12c1 and GEN-428B12c2 and the entire sequences of these were determined as in the foregoing examples.

As a result, the GEN-428B12c1 cDNA clone and the GEN-428B12c2 clone were found to have coding sequences differing by 12 amino acid residues at the 5' terminus, the GEN-428B12c1 cDNA clone being longer by 12 amino acid residues.

The GEN-428B12c1 cDNA sequence of the human NPIK gene contained an open reading frame of 2,487 nucleotides, as shown under SEQ ID NO:32, encoding an amino acid sequence comprising 829 amino acid residues, as shown under SEQ ID NO:31. The nucleotide sequence of the full-length cDNA clone comprised 3,324 nucleotides as shown under SEQ ID NO:33.

The estimated initiation codon was located, as

-76-

shown under SEQ ID NO:33, at nucleotides Nos. 115-117
corresponding to the second ATG triplet of the cDNA
clone. The termination codon was located at nucleotides
Nos. 2602-2604 and the polyadenylation signal (AATAAA) at
5 Nos. 3305-3310.

On the other hand, the GEN-428B12c2 cDNA
sequence of the human NPIK gene contained an open reading
frame of 2,451 nucleotides, as shown under SEQ ID NO:29.
The amino acid sequence encoded thereby comprised 817
10 amino acid residues, as shown under SEQ ID NO:28. The
nucleotide sequence of the full-length cDNA clone
comprised 3,602 nucleotides, as shown under SEQ ID NO:30.

The estimated initiation codon was located, as
shown under SEQ ID NO:30, at nucleotides Nos. 429-431
15 corresponding to the 7th ATG triplet of the cDNA clone.
The termination codon was located at nucleotides Nos.
2880-2882 and the polyadenylation signal (AATAAA) at Nos.
3583-3588.

(2) Northern blot analysis

20 Northern blot analysis was carried out as des-
cribed in Example 1 (2). Thus, the entire sequence of
human NPIK was amplified by PCR, the PCR product was
purified and labeled with [³²P]-dCTP (random-primed DNA
labeling kit, Boehringer Mannheim), and normal human
25 tissues were examined for expression of the human NPIK

-77-

mRNA using the MTN blot membrane with the labeled product as a probe.

As a result, the expression of the human NPIK gene was observed in 16 various human adult tissues examined and an about 3.8 kb transcript and an about 5 kb one could be detected.

Using primer A having the nucleotide sequence shown below in Table 8 and containing the initiation codon of the GEN-428B12c2 cDNA and primer B shown in table 8 and containing the termination codon, PCR was performed with Human Fetal Brain Marathon-ReadyTM cDNA (Clontech) as a template, and the nucleotide sequence of the PCR product was determined.

Table 8

15	Primer	Nucleotide sequence
	Primer A	5'-ATGGGAGATACAGTAGTGGAGC-3'
20	Primer B	5'-TCACATGATGCCGTTGGTGAG-3'

As a result, it was found that the human NPIK mRNA expressed included one lacking in nucleotides Nos. 1060-1104 of the GEN-428B12c1 cDNA sequence (SEQ ID NO:33) (amino acids Nos. 316-330 of the amino acid sequence under SEQ ID NO:31) and one lacking in nucleotides Nos. 1897-1911 of the GEN-428B12c1 cDNA sequence (SEQ ID NO:33) (amino acids Nos. 595-599 of the amino acid sequence under SEQ ID NO:31).

-78-

It was further revealed that polymorphism existed in this gene (428B12c1.fasta), as shown below in Table 9, in the region of bases Nos. 1941-1966 of the GEN-428B12c1 cDNA sequence shown under SEQ ID NO:33, whereby a mutant protein was encoded which resulted from the mutation of IQDSCEITT (amino acid residues Nos. 610-618 in the amino acid sequence (SEQ ID NO:31) encoded by GEN-428B12c1) into YKILVISA.

Table 9

			1930	1940	1950	1959
			TGGATCAAGCCAATACAAGATTCTTGTA			
			TCCATTTGGGAACAGGAGCGAGTGCCCTTTGGATCAAGCC-ATACAAGATTCTTGTA--			
1900	1910	1920	1930	1940	1950	
1960	1970	1980				
			ATTACGACTGATAGTGGCATG			
			ATTTCGGCTGATAGTGGCATGATTGAACCAAGTGGTCAATGCTGTGTCCATCCATCAGGTG			
1960	1970	1980	1990	2000	2010	

10 (3) Chromosomal mapping of human NPIK gene by FISH

Chromosomal mapping of the human NPIK gene was carried out by FISH as described in Example 1 (3).

As a result, it was found that the locus of the human NPIK gene is in the chromosomal position 1q21.1-15 q21.3.

The human NPIK gene, a novel human gene, of the present invention included two cDNAs differing in the 5' region and capable of encoding 829 and 817 amino acid

-79-

residues, as mentioned above. In view of this and further in view of the findings that the mRNA corresponding to this gene includes two deletable sites and there occurs polymorphism in a specific region corresponding to amino acid residues Nos. 610-618 of the GEN-428B12c1 amino acid sequence (SEQ ID NO:31), whereby a mutant protein is encoded, it is conceivable that human NPIK includes species resulting from a certain number of combinations, namely human NPIK, deletion-containing human NPIK, human NPIK mutant and/or deletion-containing human NPIK mutant.

Recently, several proteins belonging to the family including the above-mentioned PI3 and 4 kinases have protein kinase activity [Dhand, R., et al., EMBO J., 13, 522-533 (1994); Stack, J. H. and Emr, S. D., J. Biol. Chem., 269, 31552-31562 (1994); Hartley, K. O., et al., Cell, 82, 848-856 (1995)].

It was also revealed that a protein belonging to this family is involved in DNA repair [Hartley, K. O., et al., Cell, 82, 849-856 (1995)] and is a causative gene of ataxia [Savitsky, K., et al., Science, 268, 1749-1753 (1995)].

It can be anticipated that the human NPIK gene-encoded protein highly homologous to the family of these PI kinases is a novel enzyme phosphorylating lipids or

-80-

proteins.

According to this example, the novel human NPIK gene is provided. The use of said gene makes it possible to detect the expression of said gene in various tissues and manufacture the human NPIK protein by the technology of genetic engineering and, through these, it becomes possible to study lipid- or protein-phosphorylating enzymes such as mentioned above, study DNA repairing, study or diagnose diseases in which these are involved, for example cancer, and screen out and evaluate drugs for the treatment or prevention thereof.

(4) Construction of an expression vector for fusion protein

To subclone the coding region for a human NPIK gene (GEN-428B12c2), first of all, two primers, C1 and C2, having the sequences shown below in Table 10 were formed based on the information on the DNA sequences obtained above in (1).

Table 10

Primer	Nucleotide sequence
Primer C1	5'-CTCAGATCTATGGGAGATACAGTAGTGGAGC-3'
Primer C2	5'-TCGAGATCTTCACATGATGCCGTTGGTGAG-3'

Both of the primers C1 and C2 have a BglII site, and primer C2 is an antisense primer.

Using these two primers, cDNA derived from

-81-

human fetal brain mRNA was amplified by PCR to provide a product having a length of about 2500 bases. The amplified cDNA was precipitated from ethanol and inserted into pT7BlueT-Vector (product of Novagen) and subcloning was completed. The entire sequence was determined in the same manner as above in Examples. As a result, it was revealed that this gene had polymorphism shown above in Table 9.

The above cDNA was cleaved by BglII and subjected to agarose gel electrophoresis. The cDNA was then excised from agarose gel and collected using GENECLAN II KIT (product of Bio 101). The cDNA was inserted into pBlueBacHis2B-Vector (product of Invitrogen) at the BglII cleavage site and subcloning was completed.

The fusion vector thus obtained had a BglII cleavage site and was an expression vector for a fusion protein of the contemplated gene product (about 91 kd) and 38 amino acids derived from pBlueBacHis2B-Vector and containing a polyhistidine region and an epitope recognizing Anti-XpressTM antibody (product of Invitrogen).

(5) Transfection into insect cell Sf-9

The human NPIK gene was expressed according to the Baculovirus expression system. Baculovirus is a

-82-

cyclic double-stranded insect-pathogenic virus and can produce large amounts of inclusion bodies named polyhedrins in the cells of insects. Using Bac-N-BlueTM Transfection Kit utilizing this characteristic of Baculovirus and developed by Invitrogen, the Baculovirus expression was carried out.

Stated more specifically, 4 µg of pBlueBacHis2B containing the region of the human NPIK gene and 1 µg of Bac-N-BlueTM DNA (product of Invitrogen) were co-transfected into Sf-9 cells in the presence of InsectinTM liposomes (product of Invitrogen).

Prior to co-transfection, LacZ gene was incorporated into Bac-N-BlueTM DNA, so that LacZ would be expressed only when homologous recombination took place between the Bac-N-BlueTM DNA and pBlueBacHis2B. Thus when the co-transfected Sf-9 cells were incubated on agar medium, the plaques of the virus expressing the contemplated gene were easily detected as blue plaques.

The blue plaques were excised from each agar and suspended in 400 µl of medium to disperse the virus thereon. The suspension was subjected to centrifugation to give a supernatant containing the virus. Sf-9 cells were infected with the virus again to increase the titre and to obtain a large amount of infective virus solution.

(6) Preparation of human NPIK

-83-

The expression of the contemplated human NPIK gene was confirmed three days after infection with the virus as follows.

Sf-9 cells were collected and washed with PBS.

5 The cells were boiled with a SDS-PAGE loading buffer for 5 minutes and SDS-PAGE was performed. According to the western blot technique using Anti-XpressTM as an antibody, the contemplated protein was detected at the position of its presumed molecular weight. By contrast, in the case
10 of control cells uninfected with the virus, no band corresponding to human NPIK was observed in the same test.

Stated more specifically, three days after the infection of 15 flasks (175-cm², FALCON) of semi-
15 confluent Sf-9 cells, the cells were harvested and washed with PBS, followed by resuspension in a buffer (20 mM Tris/HCl (pH 7.5), 1 mM EDTA and 1 mM DTT). The suspended cells were lysed by 4 time-sonications for 30 seconds at 4 °C with 30 seconds intervals. The sonicated
20 cells were subjected to centrifugation and the supernatant was collected. The protein in the supernatant was immunoprecipitated using an Anti-Xpress antibody and obtained as a slurry of protein A-Sepharose beads. The slurry was boiled with a SDS-PAGE loading
25 buffer for 5 minutes. SDS-PAGE was performed for

-84-

identification and quantification of NPIK. The slurry itself was subjected to the following assaying.

(7) Confirmation of PI4 Kinase activity

NPIK was expected to have the activity of incorporation phosphoric acid at the 4-position of the inositol ring of phosphatidylinositol (PI), namely, PI4 Kinase activity.

PI4 Kinase activity of NPIK was assayed according to the method of Takenawa, et al. (Yamakawa, A. and Takenawa, T., J. Biol. Chem., 263, 17555-17560 (1988)) as shown below.

First prepared was a mixture of 10 μ l of a NPIK slurry (20 mM Tris/HCl (pH 7.5), 1 mM EDTA, 1 mM DTT and 50% protein A beads), 10 μ l of a PI solution (prepared by drying 5 mg of a PI-containing commercial chloroform solution in a stream of nitrogen onto a glass tube wall, adding 1 ml of 20 mM Tris/HCl (pH 7.5) buffer and forming micelles by sonication), 10 μ l of an applied buffer (210 mM Tris/HCl (pH 7.5), 5 mM EGTA and 100 mM MgCl_2) and 10 μ l of distilled water. Thereto was added 10 μ l of an ATP solution (5 μ l of 500 μ M ATP, 4.9 μ l of distilled water and 0.1 μ l of γ - ^{32}P ATP (6000 Ci/mmol, product of NEN Co., Ltd.)). The reaction was started at 30°C and continued for 2, 5, 10 and 20 minutes. The time 10 minutes was set as incubation time because a straight-

-85-

line increase was observed around 10 minutes in incorporation of phosphoric acid into PI in the assaying process described below.

After completion of the reaction, PI was
5 fractionated by the solvent extraction method and finally re-suspended in chloroform. The suspension was developed by thin layer chromatography (TLC) and the radioactivity of the reaction product at the PI4P-position was assayed using an analyzer (trade name: Bio-Image; product of Fuji
10 Photo Film Co., Ltd.).

Fig. 1 shows the results. Fig. 1 is an analytical diagram of the results of assaying the radioactivity based on TLC as mentioned above. The right lane (2) is the fraction of Sf-9 cell cytoplasm infected
15 with the NPIK-containing virus, whereas the left lane (1) is the fraction of uninfected Sf-9 cell cytoplasm.

Also, predetermined amounts of Triton X-100 and adenosine were added to the above reaction system to check how such addition would affect the PI4 Kinase
20 activity. The PI4 Kinase activity was assayed in the same manner as above.

Fig. 2 shows the results. The results confirmed that NPIK had a typical PI4 Kinase activity accelerated by Triton X-100 and inhibited by adenosine.

-86-

Example 10

nel-related protein type 1 (NRP1) gene and nel-related protein type 2 (NRP2) gene

5 (1) Cloning and DNA sequencing of NRP1 gene and NRP2 gene

EGF-like repeats have been found in many membrane proteins and in proteins related to growth regulation and differentiation. This motif seems to be involved in protein-protein interactions.

10 Recently, a gene encoding nel, a novel peptide containing five EGF-like repeats, was cloned from a chick embryonic cDNA library [Matsushashi, S., et al., Dev. Dynamics, 203, 212-222 (1995)]. This product is considered to be a transmembrane molecule with its EGF-
15 like repeats in the extracellular domain. A 4.5 kb transcript (nel mRNA) is expressed in various tissues at the embryonic stage and exclusively in brain and retina after hatching.

Following the procedure of Example 1 (1), cDNA
20 clones were randomly selected from a human fetal brain cDNA library and subjected to sequence analysis, followed by database searching. As a result, two cDNA clones with significantly high homology to the above-mentioned nel were found and named GEN-073E07 and GEN-093E05,
25 respectively.

-87-

Since both clones were lacking in the 5' portion, 5' RACE was performed in the same manner as in Example 2 (2) to obtain the entire coding regions.

As for the primers for 5' RACE, primers having an arbitrary sequence obtained from the cDNA sequences of the above clones were synthesized while the anchor primer attached to a commercial kit was used as such.

5' RACE clones obtained from the PCR were sequenced and the sequences seemingly covering the entire coding regions of both genes were obtained. These genes were respectively named nel-related protein type 1 (NRP1) gene and nel-related protein type 2 (NRP2) gene.

The NRP1 gene contains an open reading frame of 2,430 nucleotides, as shown under SEQ ID NO:35, the amino acid sequence deduced therefrom comprises 810 amino acid residues, as shown under SEQ ID NO:34, and the nucleotide sequence of the entire cDNA clone of said NRP1 gene comprises 2,977 nucleotides, as shown under SEQ ID NO:36.

On the other hand, the NRP2 gene contains an open reading frame of 2,448 nucleotides, as shown under SEQ ID NO:38, the amino acid sequence deduced therefrom comprises 816 amino acid residues, as shown under SEQ ID NO:37, and the nucleotide sequence of the entire cDNA clone of said NRP2 gene comprises 3,198 nucleotides, as shown under SEQ ID NO:39.

-88-

Furthermore, the coding regions were amplified by RT-PCR to exclude the possibility that either of the sequences obtained was a chimeric cDNA.

The deduced NRP1 and NRP2 gene products both
5 showed highly hydrophobic N termini capable of functioning as signal peptides for membrane insertion. As compared with chick embryonic nel, they both appeared to have no hydrophobic transmembrane domain. Comparison among NRP1, NRP2 and nel with respect to the deduced
10 peptide sequences revealed that NRP2 has 80% homology on the amino acid level and is more closely related to nel than NRP1 having 50% homology. The cysteine residues in cysteine-rich domains and EGF-like repeats were found completely conserved.

15 The most remarkable difference between the NRPs and the chick protein was that the human homologs lack the putative transmembrane domain of nel. However, even in this lacking region, the nucleotide sequences of NRPs were very similar to that of nel. Furthermore, the two
20 NRPs each possessed six EGF-like repeats, whereas nel has only five.

Other unique motifs of nel as reported by Matsubashi et al. [Matsubashi, S., et al., Dev. Dynamics, 203, 212-222 (1995)] were also found in the NRPs at
25 equivalent positions. Since as mentioned above, it was

-89-

shown that the two deduced NRP peptides are not transmembrane proteins, the NRPs might be secretory proteins or proteins anchored to membranes as a result of posttranslational modification.

5 The present inventors speculate that NRPs might function as ligands by stimulating other molecules such as EGF receptors. The present inventors further found that an extra EGF-like repeat could be encoded in nel upon frame shifting of the membrane domain region of nel.

10 When paralleled and compared with NRP2 and nel, the frame-shifted amino acid sequence showed similarities over the whole range of NRP2 and of nel, suggesting that NRP2 might be a human counterpart of nel. In contrast, NRP1 is considered to be not a human counterpart of nel
15 but a homologous gene.

(2) Northern blot analysis

Northern blot analysis was carried out as described in Example 1 (2). Thus, the entire sequences of both clones cDNAs were amplified by PCR, the PCR products
20 were purified and labeled with [³²P]-dCTP (random-primed DNA labeling kit, Boehringer Mannheim) and human normal tissues were examined for NRP mRNA expression using an MTN blot with the labeled products as two probes.

Sixteen adult tissues and four human fetal
25 tissues were examined for the expression pattern of two

-90-

NRPs.

As a result of the Northern blot analysis, it was found that a 3.5 kb transcript of NRP1 was weakly expressed in fetal and adult brain and kidney. A 3.6 kb transcript of NRP2 was strongly expressed in adult and fetal brain alone, with weak expression thereof in fetal kidney as well.

This suggests that NRPs might play a brain-specific role, for example as signal molecules for growth regulation. In addition, these genes might have a particular function in kidney.

(3) Chromosomal mapping of NRP1 gene and NRP2 gene by FISH

Chromosomal mapping of the NRP1 gene and NRP2 gene was performed by FISH as described in Example 1 (3).

As a result, it was revealed that the chromosomal locus of the NRP1 gene is localized to 11p15.1-p15.2 and the chromosomal locus of the NRP2 gene to 12q13.11-q13.12.

According to the present invention, the novel human NRP1 gene and NRP2 gene are provided and the use of said genes makes it possible to detect the expression of said genes in various tissues and produce the human NRP1 and NRP2 proteins by the technology of genetic engineering. They can further be used in the study of

-91-

the brain neurotransmission system, diagnosis of various diseases related to neurotransmission in the brain, and the screening and evaluation of drugs for the treatment and prevention of such diseases. Furthermore, the possibility is suggested that these EGF domain-containing NRPs act as growth factors in brain, hence they may be useful in the diagnosis and treatment of various kinds of intracerebral tumor and effective in nerve regeneration in cases of degenerative nervous diseases.

10

Example 11

GSPT1-related protein (GSPT1-TK) gene

(1) GSPT1-TK gene cloning and DNA sequencing

The human GSPT1 gene is one of the human homologous genes of the yeast GST1 gene that encodes the GTP-binding protein essential for the G1 to S phase transition in the cell cycle. The yeast GST1 gene, first identified as a protein capable of complementing a temperature-sensitive *gst1* (G1-to-S transition) mutant of Saccharomyces cerevisiae, was isolated from a yeast genomic library [Kikuchi, Y., Shimatake, H. and Kikuchi, A., EMBO J., 7, 1175-1182 (1988)] and encoded a protein with a target site of cAMP-dependent protein kinases and a GTPase domain.

20

The human GSPT1 gene was isolated from a KB cell cDNA library by hybridization using the yeast GST1

25

-92-

gene as a probe [Hoshino, S., Miyazawa, H., Enomoto, T.,
Hanaoka, F., Kikuchi, Y., Kikuchi, A. and Ui, M., EMBO
J., 8, 3807-3814 (1989)]. The deduced protein of said
GSPT1 gene, like yeast GST1, has a GTP-binding domain and
5 a GTPase activity center, and plays an important role in
cell proliferation.

Furthermore, a breakpoint for chromosome re-
arrangement has been observed in the GSPT1 gene located
in the chromosomal locus 16p13.3 in patients with acute
10 nonlymphocytic leukemia (ANLL) [Ozawa, K., Murakami, Y.,
Eki, T., Yokoyama, K. Soeda, E., Hoshino, S. Ui, M. and
Hanaoka, F., Somatic Cell and Molecular Genet., 18, 189-
194 (1992)].

cDNA clones were randomly selected from a human
15 fetal brain cDNA library and subjected to sequence
analysis as described in Example 1 (1) and database
searching was performed and, as a result, a clone having
a 0.3 kb cDNA sequence highly homologous to the above-
mentioned GSPT1 gene was found and named GEN-077A09. The
20 GEN-077A09 clone seemed to be lacking in the 5' region,
so that 5' RACE was carried out in the same manner as in
Example 2 (2) to obtain the entire coding region.

The primers used for the 5' RACE were P1 and P2
primers respectively having the nucleotide sequences
25 shown in Table 11 as designed based on the known cDNA

-93-

sequence of the above-mentioned cDNA, and the anchor primer used was the one attached to the commercial kit. Thirtyfive cycles of PCR were performed under the following conditions: 94°C for 45 seconds, 58°C for 45
 5 seconds and 72°C for 2 minutes. Finally, elongation reaction was carried out at 72°C for 7 minutes.

Table 11

10	Primer	Nucleotide sequence
	P1 primer	5'-GATTGTGCTCAATAATCACTATCTGAA-3'
	P2 primer	5'-GGTTACTAGGATCACAAAGTATGAATTCTGGAA-3'

15 Several of the 5' RACE clones obtained from the above PCR were sequenced and the base sequence of that cDNA clone showing overlapping between the 5' RACE clones and the GEN-077A09 clone was determined to thereby reveal the sequence regarded as covering the entire coding
 20 region. This was named GSPT1-related protein "GSPT1-TK gene".

 The GSPT1-TK gene was found to contain an open reading frame of 1,497 nucleotides, as shown under SEQ ID NO:41. The amino acid sequence deduced therefrom
 25 contained 499 amino acid residues, as shown under SEQ ID NO:40.

 The nucleotide sequence of the whole cDNA clone of the GSPT1-TK gene was found to comprise 2,057 nucleotides, as shown under SEQ ID NO:42, and the

-94-

molecular weight was calculated at 55,740 daltons.

The first methionine code (ATG) in the open reading frame had no in-frame termination codon but this ATG was surrounded by a sequence similar to the Kozak consensus sequence for translational initiation. Therefore, it was concluded that this ATG triplet occurring in positions 144-146 of the relevant sequence is the initiation codon.

Furthermore, a polyadenylation signal, AATAAA, was observed 13 nucleotides upstream from the polyadenylation site.

Human GSPT1-TK contains a glutamic acid rich region near the N terminus, and 18 of 20 glutamic acid residues occurring in this region of human GSPT1-TK are conserved and align perfectly with those of the human GSPT1 protein. Several regions (G1, G2, G3, G4 and G5) of GTP-binding proteins that are responsible for guanine nucleotide binding and hydrolysis were found conserved in the GSPT1-TK protein just as in the human GSPT1 protein.

Thus, the DNA sequence of human GSPT1-TK was found 89.4% identical, and the amino acid sequence deduced therefrom 92.4% identical, with the corresponding sequence of human GSPT1 which supposedly plays an important role in the G1 to S phase transition in the cell cycle. Said amino acid sequence showed 50.8%

-95-

identity with that of yeast GST1.

(2) Northern blot analysis

Northern blot analysis was carried out as described in Example 1 (2). Thus, the GEN-077A09 cDNA clone
5 was amplified by PCR, the PCR product was purified and labeled with [³²P]-dCTP (random-primed DNA labeling kit, Boehringer Mannheim), and normal human tissues were examined for the expression of GSPT1-TK mRNA therein using an MTN blot with the labeled product as a probe.

10 As a result of the Northern blot analysis, a 2.7 kb major transcript was detected in various tissues. The level of human GSPT1-TK expression seemed highest in brain and in testis.

(3) Chromosome mapping of GSPT1-TK gene by FISH

15 Chromosome mapping of the GSPT1-TK gene was performed by FISH as described in Example 1 (3).

As a result, it was found that the GSPT1-TK gene is localized at the chromosomal locus 19p13.3. In this chromosomal localization site, reciprocal location
20 has been observed very frequently in cases of acute lymphocytic leukemia (ALL) and acute myeloid leukemia (AML). In addition, it is reported that acute non-lymphocytic leukemia (ANLL) is associated with re-arrangements involving the human GSPT1 region [Ozawa, K.,
25 Murakami, Y., Eki, T., Yokoyama, K., Soeda, E., Hoshino,

-96-

S., Ui, M. and Hanaoka, F., Somatic Cell and Molecular Genet., 18, 189-194 (1992)].

In view of the above, it is suggested that this gene is the best candidate gene associated with ALL and
5 AML.

In accordance with the present invention, the novel human GSPT1-TK gene is provided and the use of said gene makes it possible to detect the expression of said gene in various tissues and produce the human GSPT1-TK
10 protein by the technology of genetic engineering. These can be used in the studies of cell proliferation, as mentioned above, and further make it possible to diagnose various diseases associated with the chromosomal locus of this gene, for example acute myelocytic leukemia. This
15 is because translocation of this gene may result in decomposition of the GSPT1-TK gene and further some or other fused protein expressed upon said translocation may cause such diseases.

Furthermore, it is expected that diagnosis and
20 treatment of said diseases can be made possible by producing antibodies to such fused protein, revealing the intracellular localization of said protein and examining its expression specific to said diseases. Therefore, it is also expected that the use of the gene of the present
25 invention makes it possible to screen out and evaluate

-97-

drugs for the treatment and prevention of said diseases.

DEMANDES OU BREVETS VOLUMINEUX

**LA PRÉSENTE PARTIE DE CETTE DEMANDE OU CE BREVETS
COMPREND PLUS D'UN TOME.**

CECI EST LE TOME __1__ DE __2__

NOTE: Pour les tomes additionels, veuillez contacter le Bureau Canadien des Brevets.

JUMBO APPLICATIONS / PATENTS

**THIS SECTION OF THE APPLICATION / PATENT CONTAINS MORE
THAN ONE VOLUME.**

THIS IS VOLUME __1__ OF __2__

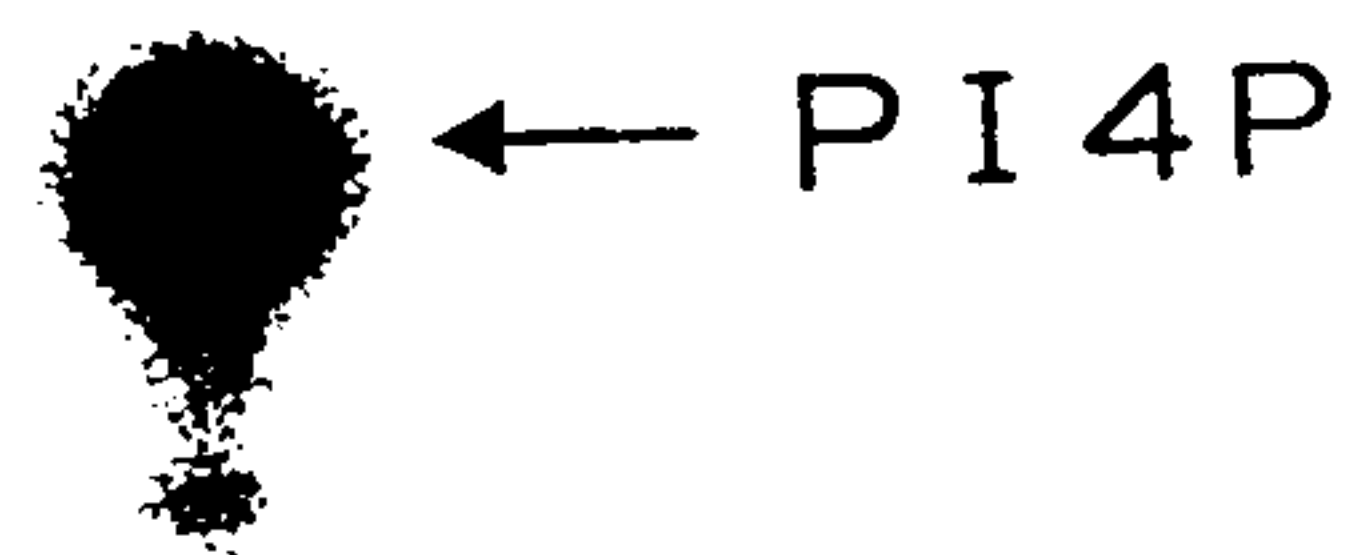
NOTE: For additional volumes please contact the Canadian Patent Office.

The embodiments of the invention in which an exclusive property or privilege is claimed are defined as follows:

1. An isolated polynucleotide encoding a peptide consisting of the amino acid sequence of SEQ ID NO:34.
2. An isolated polynucleotide comprising the DNA sequence shown in SEQ ID NO:35 or a sequence complementary thereto.
3. An isolated polynucleotide according to claim 2, which consists of the DNA sequence of SEQ ID NO:36.
4. A recombinant peptide consisting of the amino acid sequence of SEQ ID NO:34.
5. An expression vector comprising the isolated polynucleotide as defined in claim 2.
6. A transformant comprising the isolated polynucleotide as defined in claim 2.
7. A method of producing a peptide consisting of the amino acid sequence of SEQ ID NO:34 comprising culturing the transformant as defined in claim 6 and harvesting the peptide from the resulting culture.
8. An antibody specifically binding to the recombinant peptide as defined in claim 4.
9. An isolated polynucleotide encoding a peptide consisting of the amino acid sequence of SEQ ID NO:37.

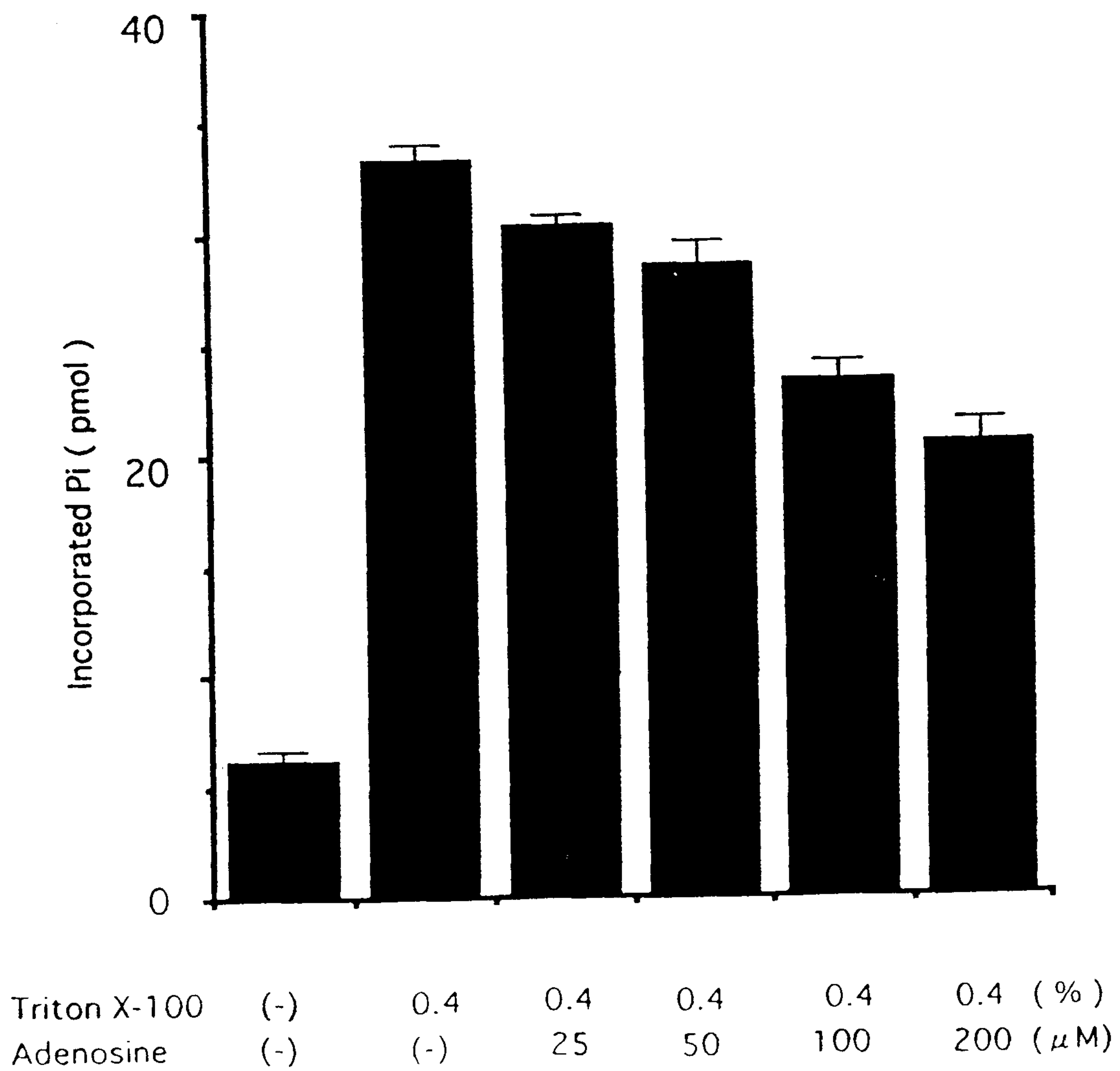
10. An isolated polynucleotide comprising the DNA sequence shown in SEQ ID NO:38 or a sequence complementary thereto.
11. An isolated polynucleotide according to claim 10, which consists of the DNA sequence of SEQ ID NO:38.
12. A recombinant peptide consisting of the amino acid sequence of SEQ ID NO:37.
13. An expression vector comprising the isolated polynucleotide as defined in claim 9.
14. A transformant comprising the isolated polynucleotide as defined in claim 9.
15. A method of producing a peptide consisting of the amino acid sequence of SEQ ID NO:37 comprising culturing the transformant of claim 14 and harvesting the peptide from the resulting culture.
16. An antibody specifically binding to the recombinant peptide as defined in claim 12.

FIG. 1



1 2

FIG. 2





P I 4 P

1

2